

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2005, 10:27:07 ; Search time 652 Seconds
(without alignments)
8734.342 Million cell updates/sec

Title: US-10-800-161-24

Perfect score: 962

Sequence: 1 tgggtttctcagaataagca.....aacattttcttcgtgtaac 962

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq_16Dec04:*

- 1: Geneseqn1980s:*
- 2: Geneseqn1990s:*
- 3: Geneseqn2000s:*
- 4: Geneseqn2001as:*
- 5: Geneseqn2001bs:*
- 6: Geneseqn2002as:*
- 7: Geneseqn2002bs:*
- 8: Geneseqn2003as:*
- 9: Geneseqn2003bs:*
- 10: Geneseqn2003cs:*
- 11: Geneseqn2003ds:*
- 12: Geneseqn2004as:*
- 13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	962	100.0	962	13	Adri16843 Arabidops
2	862	89.6	862	13	Adri16844 Arabidops
3	862	89.6	1700	12	Adri17259 Arabidops
4	862	89.6	1700	13	Adri16822 Arabidops
5	544	56.5	544	13	Adri16846 Arabidops
6	274	28.5	274	13	Adri16845 Arabidops
7	272.4	28.3	274	13	Adri16847 Arabidops
8	86	8.9	537	13	Acn53339 Cotton an
9	76.8	8.0	7657	4	Aas45477 Chemical
10	76.8	8.0	7657	6	Abi134022 Human imm
11	76.8	8.0	14006	6	Abi133958 Human imm
12	76.6	8.0	529	13	Acn56029 Cotton an
13	76.4	7.9	6644	2	Aax33181 Base sequ
14	76.4	7.9	7372	2	Aax33182 Base sequ
15	76.4	7.9	7797	2	Aax33180 Cowpox vi
16	76.4	7.9	7996	2	Aax33184 Base sequ
17	76	7.9	19124	2	Aat72882 Plasmodiu
18	76	7.9	19124	3	Aaz98287 Plasmodiu
19	75.6	7.9	700	10	Acn92384 Human col
20	74.8	7.8	9539	4	Aas45347 Chemical

C	21	74.8	7.8	9539	6	ABK28180	Abk28180 DNA trans
	22	74.4	7.7	627	13	ACN54555	Acn54555 Cotton an
	23	74.4	7.7	2520	12	ADJ48201	Adj48201 Maize oil
	24	74.2	7.7	517	13	ACN56273	Acn56273 Cotton an
	25	73.8	7.7	529	8	ABX52180	Abx52180 Bovine ES
	26	73.4	7.6	113515	6	ABL34174	Abi134174 Human imm
	27	73.2	7.6	478	13	ACN51993	Acn51993 Cotton an
	28	72.6	7.5	16033	6	ABL33404	Abi133404 Human imm
	29	72.2	7.5	778	6	ABQ15588	Abq15588 Oligonuc
	30	72.2	7.5	778	6	ABQ15589	Abq15589 Oligonuc
	31	72	7.5	446	8	ABX38235	Abx38235 Bovine ES
	32	72	7.5	16167	6	ABL33083	Abi133083 Human imm
	33	72	7.5	16167	6	ABL34529	Abi134529 Human met
	34	72	7.5	16167	6	ABL70254	Abi170254 Chemical
	35	72	7.5	16167	7	ADL99790	Adl99790 Complem
	36	71.8	7.5	34571	13	ADR66967	Adr66967 Mouse can
	37	71.6	7.4	17934	6	ABL33719	Abi133719 Human imm
	38	71.6	7.4	110000	10	ABQ84281	Abq84281 Continuation (2 of
	39	71.6	7.4	110000	10	ABQ84281	Abq84281 Continuation (3 of
	40	71.4	7.4	3505	12	ADJ48200	Adj48200 Maize oil
	41	71.2	7.4	10620	2	AAX02996	Aax02996 Human IL-
	42	71.2	7.4	14690	2	AAX22303	Aax22303 Human IL-
	43	71	7.4	447	13	ACN54792	Acn54792 Cotton an
	44	71	7.4	15548	6	ABL34155	Abi134155 Human imm
	45	70.8	7.4	3683	8	ABZ10199	Abz10199 Haematopo

ALIGNMENTS

RESULT 1

ADRI16843
ID ADRI16843 standard; DNA; 962 BP.

AC ADRI16843;

XX 21-OCT-2004 (first entry)

XX Arabidopsis thaliana N16 promoter DNA #1.

XX N16; transgenic; herbicide resistance; mouse-ear cross; promoter; ds.

XX Arabidopsis thaliana.

XX US2004154051-A1.

XX 05-AUG-2004.

XX 12-MAR-2004; 2004US-00800161.

XX 15-DEC-1999; 99US-0171008P.

XX 11-JAN-2000; 2000US-0175519P.

XX 08-DEC-2000; 2000US-00733685.

XX 20-JAN-2004; 2004US-00760752.

(SYGN) SYNGENTA PARTICIPATIONS AG.

XX Cade RM, Dietrich RA, Lawton KA;

XX WPI; 2004-580223/56.

XX Novel isolated nucleic acid molecule comprising Arabidopsis N16 promoter, useful in regulating transcription of coding sequence of interest.

XX Claim 11; SEQ ID NO 24; 39pp; English.

XX The present invention relates to an Arabidopsis N16 gene promoter useful in regulating transcription of coding sequence of interest. The invention is useful in production of transgenic plant or seed exhibiting resistance to herbicides and microorganism such as Phytophthora parasitica, Pseudomonas syringae, Cercospora nicotianae, Peronospora parasitica. The present sequence is Arabidopsis thaliana N16 promoter DNA. This sequence

CC is used in the exemplification of the invention.

XX	Sequence	962 BP; 379 A; 125 C; 165 G; 293 T; 0 U; 0 Other;	
SQ	Query Match	100.0%; Score 962; DB 13; Length 962;	
	Best Local Similarity	100.0%; Pred. No. 1.8e-164;	
	Matches	962; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1	TGTGTTTCTCAGAAATAGACGAAATATTATATAAAGCATGCAATCTCTTATAGATCG 60	
Db	1	TGTGTTTCTCAGAAATAGACGAAATATTATATAAAGCATGCAATCTCTTATAGATCG 60	
QY	61	CGAAGTTTAAAAAACAATAGAAATGTTACATATTTACATGGTTTTTATTGGATAACA 120	
Db	61	CGAAGTTTAAAAAACAATAGAAATGTTACATATTTACATGGTTTTTATTGGATAACA 120	
QY	121	TGACAAATATTTATTTATTTTCATGAGTTTTTATTGGATAGCATGACAAATATTAATATAT 180	
Db	121	TGACAAATATTTATTTATTTTCATGAGTTTTTATTGGATAGCATGACAAATATTAATATAT 180	
QY	181	CAGTGTAAATAACATGTTTGTTCCTTAAATACATGCAATTTAAATCAGACATTTGTTT 240	
Db	181	CAGTGTAAATAACATGTTTGTTCCTTAAATACATGCAATTTAAATCAGACATTTGTTT 240	
QY	241	TAAATCAAATCTAATCTCTTATATCACACGACATTCGCGAAATTCAGGTAAAAAGA 300	
Db	241	TAAATCAAATCTAATCTCTTATATCACACGACATTCGCGAAATTCAGGTAAAAAGA 300	
QY	301	GAAATATAAGAAATGAGAGATAGAGATTTCTATGGAAAAAGAAAGAGACATGTAGG 360	
Db	301	GAAATATAAGAAATGAGAGATAGAGATTTCTATGGAAAAAGAAAGAGACATGTAGG 360	
QY	361	TGAACAAAATAAAGAGATATGATGATATATTTTATGAGAGTGGTGAAGATTTTATAG 420	
Db	361	TGAACAAAATAAAGAGATATGATGATATATTTTATGAGAGTGGTGAAGATTTTATAG 420	
QY	421	AGAGGAGAGAGAAATAGAAAAAGAAATGACATGGTGAATCTGAAGAGATGAATCTG 480	
Db	421	AGAGGAGAGAGAAATAGAAAAAGAAATGACATGGTGAATCTGAAGAGATGAATCTG 480	
QY	481	TTAAAGATGAAGAGAGAAAGAGAACTCCATGGCTAAAGTCTCGTAAAGAAAGTGAAG 540	
Db	481	TTAAAGATGAAGAGAGAAAGAGAACTCCATGGCTAAAGTCTCGTAAAGAAAGTGAAG 540	
QY	541	AAACAAAGAGAAAGAGAAAGAGAAAGCTTAAATAGACTAACTATTGCCAAAATTC 600	
Db	541	AAACAAAGAGAAAGAGAAAGAGAAAGCTTAAATAGACTAACTATTGCCAAAATTC 600	
QY	601	TGTAGCGGACAAATACATATTTTGGTCCAGGTTATTTGTGATTTCTTTGAAAGTCAA 660	
Db	601	TGTAGCGGACAAATACATATTTTGGTCCAGGTTATTTGTGATTTCTTTGAAAGTCAA 660	
QY	661	TTATTTCTTACATATATCTCTTAAAAATATAGCCGATACCAATTTTCCACACATGGACTTC 720	
Db	661	TTATTTCTTACATATATCTCTTAAAAATATAGCCGATACCAATTTTCCACACATGGACTTC 720	
QY	721	CTTTATTTCCAAAAGTCAATAAGTGTGACGTGATGATCTAGCTTTTAAAAACATCGCAT 780	
Db	721	CTTTATTTCCAAAAGTCAATAAGTGTGACGTGATGATCTAGCTTTTAAAAACATCGCAT 780	
QY	781	GATGATGTCATTTAGCATCAATCTCCACCGTCCCAATTTATTTAGTTGTTGACAAATTCGAC 840	
Db	781	GATGATGTCATTTAGCATCAATCTCCACCGTCCCAATTTATTTAGTTGTTGACAAATTCGAC 840	
QY	841	CGTCTAAGTTCCACACCGCGCTATAAGAGTTTCAATATATAAATTTTAGCAAAATAAAT 900	
Db	841	CGTCTAAGTTCCACACCGCGCTATAAGAGTTTCAATATATAAATTTTAGCAAAATAAAT 900	
QY	901	CAGCAAAATAATTTTTCTTGACTTAAGCTTAAACGACGCCGTTAAACATTTTCTCTGGCTA 960	
Db	901	CAGCAAAATAATTTTTCTTGACTTAAGCTTAAACGACGCCGTTAAACATTTTCTCTGGCTA 960	
QY	961	AC 962	

Db	961	AC 962	
RESULT 2			
ADRI6844			
ID	ADRI6844	standard; DNA; 862 BP.	
XX	ADRI6844;		
XX	AC		
XX	21-OCT-2004	(first entry)	
DE	Arabidopsis thaliana	NI16; transgenic; herbicide resistance; mouse-ear cress; promoter; ds.	
XX	Arabidopsis thaliana.		
XX	US2004154051-A1.		
XX	05-AUG-2004.		
XX	12-MAR-2004; 2004US-00800161.		
XX	15-DEC-1999; 99US-0171008P.		
PR	11-JAN-2000; 2000US-0175519P.		
PR	08-DEC-2000; 2000US-00733685.		
PR	20-JAN-2004; 2004US-00760752.		
XX	(SYGN) SYNGENTA PARTICIPATIONS AG.		
PA	Cade RM, Dietrich RA, Lawton KA;		
PI	WPI; 2004-580223/56.		
DR	Novel isolated nucleic acid molecule comprising Arabidopsis NI16 promoter, useful in regulating transcription of coding sequence of interest.		
PT	Claim 12; SEQ ID NO 25; 39pp; English.		
XX	The present invention relates to an Arabidopsis NI16 gene promoter useful in regulating transcription of coding sequence of interest. The invention is useful in production of transgenic plant or seed exhibiting resistance to herbicides and microorganism such as Phytophthora parasitica, Pseudomonas syringae, Cercospora nicotianae, Peronospora parasitica. The present sequence is Arabidopsis thaliana NI16 promoter DNA. This sequence is used in the exemplification of the invention.		
CC	Sequence 862 BP; 337 A; 112 C; 152 G; 261 T; 0 U; 0 Other;		
SQ	Query Match	89.6%; Score 862; DB 13; Length 862;	
	Best Local Similarity	100.0%; Pred. No. 2e-146;	
	Matches	862; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	101	TGGGTTTTTATTGGATAACATGACAAAATATTTATTATTTCATGAGTTTTTATTGGATAG 160	
Db	1	TGGGTTTTTATTGGATAACATGACAAAATATTTATTATTTCATGAGTTTTTATTGGATAG 60	
QY	161	CATGACAAATATTAAATATATATCATGTTTAATAACATGTTTGTCTTAAAAATACATGCATT 220	
Db	61	CATGACAAATATTAAATATATATCATGTTTAATAACATGTTTGTCTTAAAAATACATGCATT 120	
QY	221	TTAAAAATCAGACATTTGTTTTTAAAAATCAAATCTTAATCTCTTATATCACACGACATTGAC 280	
Db	121	TTAAAAATCAGACATTTGTTTTTAAAAATCAAATCTTAATCTCTTATATCACACGACATTGAC 180	
QY	281	GGAAAAATTCAGGTAAAAAGAGAAAAATAAAGATGAGAGATAGAGAGATTTCTTATGAAAA 340	
Db	181	GGAAAAATTCAGGTAAAAAGAGAAAAATAAAGATGAGAGATAGAGAGATTTCTTATGAAAA 240	
QY	341	AGAAAGAGAGAAACATGTCAGTGAACAAAAATAAAGAGATATGATATATTTTATCAGAG 400	

```
Db 241 AGAAGAGAGACATGTAGTGGAACAAATAAAGAGATATGATGATATATTTTATGAGAG 300
Qy 401 GTGCTGAAGATATTTTAGGAGAGGAGAGAAATAGAAAAGAAATGACATGGTGAA 460
Db 301 GTGCTGAAGATATTTTAGGAGAGGAGAGAAATAGAAAAGAAATGACATGGTGAA 360
Qy 461 TCTGAAGAGATGAATTTGTGTTAAAGATGAAGAGAGAAAGAACTCATGGCTTAAAGTC 520
Db 361 TCTGAAGAGATGAATTTGTGTTAAAGATGAAGAGAGAAAGAACTCATGGCTTAAAGTC 420
Qy 521 TCGTAAAGAGATCAAAAAGAAACAAAGAGAGGAGAGAAAGAAAGGCTTAAATAGA 580
Db 421 TCGTAAAGAGATCAAAAAGAAACAAAGAGAGGAGAGAAAGAAAGGCTTAAATAGA 480
Qy 581 CTAACATATGCCAAAATTTCTGTAGCCGACAAATACTATTTGGTCCAAAGGTTATTTGTG 640
Db 481 CTAACATATGCCAAAATTTCTGTAGCCGACAAATACTATTTGGTCCAAAGGTTATTTGTG 540
Qy 641 TATCTTTTGAAGTCAAAAGTTATTTCTTACATATACTCTTAAATAATATAGCCGATACAA 700
Db 541 TATCTTTTGAAGTCAAAAGTTATTTCTTACATATACTCTTAAATAATATAGCCGATACAA 600
Qy 701 TTTTCCACACATCGATCTCTTTATTTCCAAAAGTCAATAAAGTGTGACGTCATGATACT 760
Db 601 TTTTCCACACATCGATCTCTTTATTTCCAAAAGTCAATAAAGTGTGACGTCATGATACT 660
Qy 761 TAGCTTTTAAACATCGCATGATGTCATTAGCATCAATCTCCACCGTCCCAATTTATT 820
Db 661 TAGCTTTTAAACATCGCATGATGTCATTAGCATCAATCTCCACCGTCCCAATTTATT 720
Qy 821 TAGTTGTTGACATATCGACCGTCTAAGTTCCACACCGCGCTTAAAGAGTTTCATTAT 880
Db 721 TAGTTGTTGACATATCGACCGTCTAAGTTCCACACCGCGCTTAAAGAGTTTCATTAT 780
Qy 881 AAATTTTACCAATAAATAACAGCAATAATTTTCTTGACTTAAGCTTAAACGACCCG 940
Db 781 AAATTTTACCAATAAATAACAGCAATAATTTTCTTGACTTAAGCTTAAACGACCCG 840
Qy 941 TTAACATTTTCTTCTGGCTAAC 962
Db 841 TTAACATTTTCTTCTGGCTAAC 862
```

RESULT 3

```
ADN17259
ID ADN17259 standard; DNA; 1700 BP.
XX
AC ADN17259;
XX
XX 17-JUN-2004 (first entry)
XX Arabidopsis thaliana Nil6 gene.
XX Nil6; SAR; systemic acquired resistance; mouse-ear cross; gene; db.
XX Arabidopsis thaliana.
PH Key Location/Qualifiers
FT misc_feature 365..374
FT /*tag= a
FT /*note= "TC1 motif"
FT misc_feature 426..435
FT /*tag= b
FT /*note= "TC1 motif"
FT misc_feature 609..614
FT /*tag= c
FT /*note= "MYCATR22 element"
FT misc_feature 646..665
FT /*tag= d
FT /*note= "CMV AS1 salicylic acid response element"
FT misc_feature 707..712
FT /*tag= e
FT /*note= "PAL box"
```

```
FT misc_feature 757..762
FT /*tag= f
FT /*note= "HEXAMERAT 4 element"
FT CDS 863..1231
FT /*tag= g
FT /*product= "Arabidopsis thaliana Nil6 protein"
XX
PN US6706952-B1.
XX
PD 16-MAR-2004.
XX
XX 08-DEC-2000; 2000US-00733685.
XX
XX 15-DEC-1999; 99US-0171008P.
XX 11-JAN-2000; 2000US-0175519P.
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
XX Cad RM, Dietrich RA;
XX
XX WPI; 2004-313378/29.
XX P-PSDB; ADN17258.
XX
XX New Nil6 nucleic acid sequence and encoded protein, useful for increasing
XX systemic acquired resistance gene expression in a plant.
XX
XX Example 2; SEQ ID NO 3; 29pp; English.
XX
XX The invention relates to Arabidopsis Nil6 gene encoding a protein
XX involved in the regulation of SAR gene expression in plants. The Nil6
XX nucleic acid molecule and the encoded protein is useful in increasing
XX systemic acquired resistance (SAR) gene expression in a plant. The
XX present sequence is Arabidopsis thaliana Nil6 gene.
XX
XX Sequence 1700 BP; 593 A; 237 C; 361 G; 509 T; 0 U; 0 Other;
XX
Query Match 89.6%; Score 862; DB 12; Length 1700;
Best Local Similarity 100.0%; Pred. No. 2.1e-146;
Matches 862; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 101 TGGGTTTTTATGGAATAACATGACAAATATTTATTTATTTTCATGAGTTTATTTGGATAG 160
Db 1 TGGGTTTTTATGGAATAACATGACAAATATTTATTTATTTTCATGAGTTTATTTGGATAG 60
Qy 161 CATGACAAATATTAATATATATCATGTTTAATAACATGTTTGTCTTAAATACATGCAAT 220
Db 61 CATGACAAATATTAATATATATCATGTTTAATAACATGTTTGTCTTAAATACATGCAAT 120
Qy 221 TTAATAATCAGACATTTGTTTTTAAATCAAAATCTAATCTCTTATATCAACACGACATTGAC 280
Db 121 TTAATAATCAGACATTTGTTTTTAAATCAAAATCTAATCTCTTATATCAACACGACATTGAC 180
Qy 281 GGAATAATTCAGGTAAAAAGAGAAAAATAAGAAATGAGAGATAGAGAGATTTCTATGGAATA 340
Db 181 GGAATAATTCAGGTAAAAAGAGAAAAATAAGAAATGAGAGATAGAGAGATTTCTATGGAATA 240
Qy 341 AGAAGAGAGACATGTCAGTGAACAAATAAAGAGATATGATGATATATTTTATGAGAG 400
Db 241 AGAAGAGAGACATGTCAGTGAACAAATAAAGAGATATGATGATATATTTTATGAGAG 300
Qy 401 GTGGTGAAGATTTATTTTAGGAGAGGAGAGAAAAATAGAAAAAGAAATACATGGTGAA 460
Db 301 GTGGTGAAGATTTATTTTAGGAGAGGAGAGAAAAATAGAAAAAGAAATACATGGTGAA 360
Qy 461 TCTGAAAGAGATGAATTTGTGTTAAAGATGAAGAGAGAAAAAGAGAACTCCATGGCTTAAAGTC 520
Db 361 TCTGAAAGAGATGAATTTGTGTTAAAGATGAAGAGAGAAAAAGAGAACTCCATGGCTTAAAGTC 420
Qy 521 TCGTAAAGAGATGAATAAAGAGAAACAAAGAGAGAAAGAGAGAAAGGCTTAAATAGA 580
Db 421 TCGTAAAGAGATGAATAAAGAGAAACAAAGAGAGAAAGAGAGAAAGGCTTAAATAGA 480
Qy 581 CTAACTATTGCCAAAATTTCTTGAGCGCAAAATACTATTTGGTCCAAAGGTTATTTGTG 640
```


QY 761 TAGCTTTAAACATCCGATGATGTCATTAGCATCAATCTCCACCGTCCCAATTTATT 820
 DB 661 TAGCTTTAAACATCCGATGATGTCATTAGCATCAATCTCCACCGTCCCAATTTATT 720
 QY 821 TAGTTGTTGACAATATCGACGGTCTAAGTTCCACACGACGGCTATAAGAGTTTCATTAT 880
 DB 721 TAGTTGTTGACAATATCGACGGTCTAAGTTCCACACGACGGCTATAAGAGTTTCATTAT 780
 QY 881 AAATTTTAGCAAAATAAAATCAGCAATAATTTTCTTGACTAAGCTTAAACGACGCCG 940
 DB 781 AAATTTTAGCAAAATAAAATCAGCAATAATTTTCTTGACTAAGCTTAAACGACGCCG 840
 QY 941 TTAACATTTTCTTGCGCTAAC 962
 DB 841 TTAACATTTTCTTGCGCTAAC 862

RESULT 5
 ADR16846
 ID ADR16846 standard; DNA; 544 BP.
 XX AC
 AC ADR16846;
 XX DT 21-OCT-2004 (first entry)
 DE Arabidopsis thaliana N16 promoter DNA #4.
 XX KW N16; transgenic; herbicide resistance; mouse-ear cress; promoter; ds.
 XX OS Arabidopsis thaliana.
 XX PN US2004154051-A1.
 XX PD 05-AUG-2004.
 XX PF 12-MAR-2004; 2004US-00800161.
 XX PR 15-DEC-1999; 99US-0171008P.
 PR 11-JAN-2000; 2000US-0175519P.
 PR 08-DEC-2000; 2000US-00733685.
 PR 20-JAN-2004; 2004US-00760752.
 XX (SYGN) SYNGENTA PARTICIPATIONS AG.
 PA Cade RM, Dietrich RA, Lawton KA;
 PI WPI; 2004-580223/56.
 XX Novel isolated nucleic acid molecule comprising Arabidopsis N16 promoter, useful in regulating transcription of coding sequence of interest.
 XX Claim 12; SEQ ID NO 27; 39pp; English.
 XX The present invention relates to an Arabidopsis N16 gene promoter useful in regulating transcription of coding sequence of interest. The invention is useful in production of transgenic plant or seed exhibiting resistance to herbicides and microorganism such as *Phytophthora parasitica*, *Pseudomonas syringae*, *Cercospora nicotianae*, *Peronospora parasitica*. The present sequence is Arabidopsis thaliana N16 promoter DNA. This sequence is used in the exemplification of the invention.
 XX Sequence 544 BP; 209 A; 84 C; 95 G; 156 T; 0 U; 0 Other;
 SQ Query Match 56.5%; Score 544; DB 13; Length 544;
 Best Local Similarity 100.0%; Pred. NO. 4.5e-89;
 Matches 544; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 408 AGATTATTTTAGAGAGGGAGAGAAATAGAAAAGAAAATGACATGGTGAATCTGAAG 467
 DB 1 AGATTATTTTAGAGAGGGAGAGAAATAGAAAAGAAAATGACATGGTGAATCTGAAG 60

QY 468 AAGATGAATTTGTGTTAAAGATGAAGAGAGAGAAAGGCTCCATGGCTAAAGTCTCGTAAA 527
 DB 61 AAGATGAATTTGTGTTAAAGATGAAGAGAGAGAAAGGCTCCATGGCTAAAGTCTCGTAAA 120
 QY 528 GAAGATGAAAAGAAAGAAAG 587
 DB 121 GAAGATGAAAAGAAAGAAAG 180
 QY 588 TTGCCAAAATTTCTGTAGCCGACAAATATCTATTGTTGGTCCAAAGGTTATTTTGTGTATCTT 647
 DB 181 TTGCCAAAATTTCTGTAGCCGACAAATATCTATTGTTGGTCCAAAGGTTATTTTGTGTATCTT 240
 QY 648 TTGAAGTCAAAAGTTATTTCTTACATATCTCTAAATAATAGCCGATACCAATTTTTC 707
 DB 241 TTGAAGTCAAAAGTTATTTCTTACATATCTCTAAATAATAGCCGATACCAATTTTTC 300
 QY 708 ACACATGGGCTTCTTTATTTCCAAAAGTCAATAAAGTGTGACGTCATGATACCTTACGCTT 767
 DB 301 ACACATGGGCTTCTTTATTTCCAAAAGTCAATAAAGTGTGACGTCATGATACCTTACGCTT 360
 QY 768 TAAACATCGCATGATGATGTCATTAGCATCAATCTCCACCGTCCCAATTTTATTAGTTGT 827
 DB 361 TAAACATCGCATGATGATGTCATTAGCATCAATCTCCACCGTCCCAATTTTATTAGTTGT 420
 QY 828 TGACAATATCGACCGTCTAAGTTCCACACGCGCTATAAGAGTTTTCATTATAAATTTT 887
 DB 421 TGACAATATCGACCGTCTAAGTTCCACACGCGCTATAAGAGTTTTCATTATAAATTTT 480
 QY 888 AGCAAAATAAAATCAGCAATAATTTTCTTGACTAAGCTTAAACGACGCCGTTAAACAT 947
 DB 481 AGCAAAATAAAATCAGCAATAATTTTCTTGACTAAGCTTAAACGACGCCGTTAAACAT 540
 QY 948 TTTC 951
 DB 541 TTTC 544

RESULT 6
 ADR16845
 ID ADR16845 standard; DNA; 274 BP.
 XX AC
 AC ADR16845;
 XX DT 21-OCT-2004 (first entry)
 DE Arabidopsis thaliana N16 promoter DNA #3.
 XX KW N16; transgenic; herbicide resistance; mouse-ear cress; promoter; ds.
 XX OS Arabidopsis thaliana.
 XX PN US2004154051-A1.
 XX PD 05-AUG-2004.
 XX PF 12-MAR-2004; 2004US-00800161.
 XX PR 15-DEC-1999; 99US-0171008P.
 PR 11-JAN-2000; 2000US-0175519P.
 PR 08-DEC-2000; 2000US-00733685.
 PR 20-JAN-2004; 2004US-00760752.
 XX (SYGN) SYNGENTA PARTICIPATIONS AG.
 PA Cade RM, Dietrich RA, Lawton KA;
 PI WPI; 2004-580223/56.
 XX Novel isolated nucleic acid molecule comprising Arabidopsis N16 promoter, useful in regulating transcription of coding sequence of interest.
 XX Claim 1; SEQ ID NO 26; 39pp; English.

XX The present invention relates to an Arabidopsis N116 gene promoter useful
 CC in regulating transcription of coding sequence of interest. The invention
 CC is useful in production of transgenic plant or seed exhibiting resistance
 CC to herbicides and microorganism such as Phytophthora parasitica,
 CC Pseudomonas syringae, Cercospora nicotianae, Peronospora parasitica. The
 CC present sequence is Arabidopsis thaliana N116 promoter DNA. This sequence
 CC is used in the exemplification of the invention.
 XX SQ Sequence 274 BP; 92 A; 58 C; 35 G; 89 T; 0 U; 0 Other;

Query Match 28.5%; Score 274; DB 13; Length 274;
 Best Local Similarity 100.0%; Pred. No. 2.2e-40;
 Matches 274; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 678 TCTAAAAATATAGCCGATACCAATTTTCCACACATGGACTTCCTTTATTCCAAAGTCA 737
 Db 1 TCTAAAAATATAGCCGATACCAATTTTCCACACATGGACTTCCTTTATTCCAAAGTCA 60
 QY 738 ATAAAGTGTGACGTCATGATCTTACGCTTTAAACATCGCATGATGTCATTAGCAT 797
 Db 61 ATAAAGTGTGACGTCATGATCTTACGCTTTAAACATCGCATGATGTCATTAGCAT 120
 QY 798 CAATCTCCACCGTCCCAATTTATTAGTTGTGACAATATCGACCGCTAAAGTTCCACACC 857
 Db 121 CAATCTCCACCGTCCCAATTTATTAGTTGTGACAATATCGACCGCTAAAGTTCCACACC 180
 QY 858 GACGGCTATAAGAGTTTCATTATAAATTTTAGCAAAATAAAATCAGCAAAATAATTTTTC 917
 Db 181 GACGGCTATAAGAGTTTCATTATAAATTTTAGCAAAATAAAATCAGCAAAATAATTTTTC 240
 QY 918 TTGACTAAGCTTAAACGACCGCTTAACATTTTC 951
 Db 241 TTGACTAAGCTTAAACGACCGCTTAACATTTTC 274

RESULT 7
 ADR16847
 ID ADR16847 standard; DNA; 274 BP.
 XX AC ADR16847;
 XX DT 21-OCT-2004 (first entry)
 XX DE Arabidopsis thaliana N116 promoter DNA #5.
 XX KW N116; transgenic; herbicide resistance; mouse-ear cross; promoter; ds.
 XX OS Arabidopsis thaliana.
 XX PN US2004154051-A1.
 XX PD 05-AUG-2004.
 XX PF 12-MAR-2004; 2004US-00800161.
 XX PR 15-DEC-1999; 99US-0171008P.
 XX PR 11-JAN-2000; 2000US-0175519P.
 XX PR 08-DEC-2000; 2000US-00733685.
 XX PR 20-JAN-2004; 2004US-00760752.
 XX (SYGN) SYNGENTA PARTICIPATIONS AG.
 XX PA Cade RM, Dietrich RA, Lawton KA;
 XX PI WPI; 2004-580223/56.
 XX DR Novel isolated nucleic acid molecule comprising Arabidopsis N116
 XX PT promoter, useful in regulating transcription of coding sequence of
 XX PT interest.
 XX PS Claim 1; SEQ ID NO 28; 39pp; English.
 XX

CC The present invention relates to an Arabidopsis N116 gene promoter useful
 CC in regulating transcription of coding sequence of interest. The invention
 CC is useful in production of transgenic plant or seed exhibiting resistance
 CC to herbicides and microorganism such as Phytophthora parasitica,
 CC Pseudomonas syringae, Cercospora nicotianae, Peronospora parasitica. The
 CC present sequence is Arabidopsis thaliana N116 promoter DNA. This sequence
 CC is used in the exemplification of the invention.
 XX SQ Sequence 274 BP; 91 A; 58 C; 36 G; 89 T; 0 U; 0 Other;

Query Match 28.3%; Score 272.4; DB 13; Length 274;
 Best Local Similarity 99.6%; Pred. No. 4.3e-40;
 Matches 273; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 678 TCTAAAAATATAGCCGATACCAATTTTCCACACATGGACTTCCTTTATTCCAAAGTCA 737
 Db 1 TCTAGAAATATAGCCGATACCAATTTTCCACACATGGACTTCCTTTATTCCAAAGTCA 60
 QY 738 ATAAAGTGTGACGTCATGATCTTACGCTTTAAACATCGCATGATGTCATTAGCAT 797
 Db 61 ATAAAGTGTGACGTCATGATCTTACGCTTTAAACATCGCATGATGTCATTAGCAT 120
 QY 798 CAATCTCCACCGTCCCAATTTATTAGTTGTGACAATATCGACCGCTAAAGTTCCACACC 857
 Db 121 CAATCTCCACCGTCCCAATTTATTAGTTGTGACAATATCGACCGCTAAAGTTCCACACC 180
 QY 858 GACGGCTATAAGAGTTTCATTATAAATTTTAGCAAAATAAAATCAGCAAAATAATTTTTC 917
 Db 181 GACGGCTATAAGAGTTTCATTATAAATTTTAGCAAAATAAAATCAGCAAAATAATTTTTC 240
 QY 918 TTGACTAAGCTTAAACGACCGCTTAACATTTTC 951
 Db 241 TTGACTAAGCTTAAACGACCGCTTAACATTTTC 274

RESULT 8
 ACN53339
 ID ACN53339 standard; cDNA; 537 BP.
 XX AC ACN53339;
 XX DT 02-DEC-2004 (first entry)
 XX DE Cotton androecium tissue EST Clone ID: LIB3828-006-Q1-N6-D9, SEQ:8120.
 XX KW Cotton; plant; EST; expressed sequence tag; transgenic plant; androecium;
 KW variety Nutcoton33B; library LIB3828; molecular tag; molecular marker;
 KW genetic mapping; molecular mapping; seed germination; plant growth;
 KW plant quality; plant yield; plant breeding; tissue printing; ss.
 XX OS Gossypium hirsutum.
 XX PN US2004123340-A1.
 XX PD 24-JUN-2004.
 XX PF 12-DEC-2001; 2001US-00021323.
 XX PR 14-DEC-2000; 2000US-0255619P.
 XX (DEIK/) DEIKMAN J.
 XX PA (FENG/) FENG P C C.
 XX PA (FINC/) FINCHER K L.
 XX PA (ZIEG/) ZIEGLER T E.
 XX Deikman J, Feng PCC, Fincher KL, Ziegler TE;
 XX WPI; 2004-479808/45.
 XX DR New isolated nucleic acid molecule that encodes a plant protein or its
 XX PT fragment, useful for isolating a variety of agronomically significant
 XX PT genes associated with plant growth, quality or yield, and as molecular
 XX PT tags to map genes.

PR	30-JUN-2000; 2000DE-01032529.
PR	01-SEP-2000; 2000DE-01043826.
XX	
PA	(EPIG-) EPIGENOMICS AG.
XX	
PI	Olek A, Piepenbrock C, Berlin K;
PI	
DR	WPI; 2002-130909/17.
XX	
PT	Nucleic acid comprising fragment of chemically modified gene, useful for
PT	diagnosis and treatment of diseases associated with abnormal cytosine
PT	methylation.
XX	
PS	Claim 1; SEQ ID NO 1931; 32pp + Sequence Listing; German.
XX	
CC	The present invention provides a number of human immune system associated
CC	genes which are modified by the methylation of cytosines. The sequences
CC	can be used in the diagnosis and treatment of immune system disorders,
CC	including eye diseases such as retinopathy, neovascular glaucoma and
CC	macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC	leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC	rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC	diseases. The present sequence is a gene of the invention
XX	
SQ	Sequence 14006 BP; 3278 A; 155 C; 3257 G; 7313 T; 0 U; 3 Other;
	Query Match 8.0%; Score 76.8; DB 6; Length 14006;
	Best Local Similarity 47.7%; Pred. No. 0.00011;
	Matches 225; Conservative 0; Mismatches 247; Indels 0; Gaps 0;
Qy	125 AAATATTATTATTTCATGAGTGTTCATTCGGATGACATGACAATATTAATATATCAGT 184
Db	
	2628 AATAATAACTAAAAACAACAAACCTTTAATAACAACGACCTAATTTAAATTCCTATCGC 2569
Qy	185 GTTAATAACATGTTTGTTCCTTAAATAACATGCAATTTTAAATCAGACATTTGTTTAAA 244
Db	
	2568 TATCACTATAAACTATATAAATCTTAAACAAAACCTTTCACCTTTTAAATCTCAATTTCTC 2509
Qy	245 ATCAAACTCTAATCTCTTATATACAAACGACATTCACGGAATAATTCAGGTAAAAAGAGAAA 304
Db	
	2508 ATCTACTATAAACAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATA 2449
Qy	305 ATAAGAATGAGAGATGAGAGATTTCTTATGGAAAAAGAAAGAGAGAACATGTAGTGCAA 364
Db	
	2448 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATA 2389
Qy	365 CAAATTAAGAGATATGATGATATATTTTATGAGAGGTGGTGAAGATTTATTTTAGGAGAG 424
Db	
	2388 AAAAAAACCAATTAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATA 2329
Qy	425 GGAGAGAGAAATAGAAAAAGAAATGACATGGTGAATCTGGAAGAAGATGAATGTTGTTAA 484
Db	
	2328 ATAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATA 2269
Qy	485 AGATGAAGAGAGAAAGAGAACTCCATCGCTAAAGTCTCGTAAAGAAGATGAATAAGAAC 544
Db	
	2268 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATA 2209
Qy	545 AAAAGAAGGAGAGAAAGAGAAAGGCGTAAATAGACTAACTATTGCCAAAA 596
Db	
	2208 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATA 2157
RESULT 12	
ACN56029	
ID	ACN56029 standard; cDNA; 529 BP.
XX	
AC	ACN56029;
XX	
XX	02-DEC-2004 (first entry)
DT	
XX	Cotton androecium tissue EST Clone ID: LIB3828-031-Q6-N6-C8, SEQ:10810.
DE	


```
XX 25-JUN-1999 (first entry)
XX Base sequence of the plasmid pRx-Bcl-xl-bsr.
XX
XX Cowpox virus; bsr; viral vector; expression; apoptosis; resistance; crmA;
XX bcl-2; bcl-xl; FLIP; survivin; IAP; ILP; adenovirus; cancer;
XX autoimmune disease; graft rejection reaction; inflammation;
XX inflammatory disease; ss.
XX
XX Synthetic.
XX Homo sapiens.
XX
XX WO9913073-A2.
XX
XX 18-MAR-1999.
XX
XX 07-SEP-1998; 98WO-JP004010.
XX
XX 08-SEP-1997; 97JP-00259235.
XX
XX (RPRG-) RPR GENCELL ASIA PACIFIC INC.
XX
XX Hamada H;
XX
XX WPI; 1999-243728/20.
XX
XX New apoptosis-resistant virus-sensitive cell.
XX
XX Example 2; Page 41-45; 51pp; English.
XX
XX The present invention describes an apoptosis-resistant virus-sensitive
XX cell line into which an apoptosis resistance gene has been introduced.
XX The recombinant viruses generated are capable of expressing apoptosis-
XX associated genes. These can then be used in a variety of diseases for
XX which the induction of apoptosis by gene transfer, or where the
XX inhibition of harmful apoptosis, is therapeutic. The recombinant viruses
XX are useful as vectors for gene therapy which can be applied to cancer
XX therapy for destroying cancer cells selectively, the treatment of
XX autoimmune diseases and graft rejection reaction, and apoptosis induction
XX therapy for inflammatory cells in inflammatory diseases. Prior arts have
XX encountered the problem where if an adenovirus vector capable of
XX expressing an apoptosis-associated gene is introduced into animal cells,
XX the cells producing the virus will be destroyed because the period of
XX time required to induce cell death by apoptosis is shorter than that
XX required to replicate and produce the virus, resulting in failure to
XX obtain a recombinant virus having the integrated apoptosis-associated
XX gene. In this invention an apoptosis-resistant 293 cell line (having an
XX apoptosis resistant gene introduced) is established and overcomes the
XX problem. The present sequence represents the base sequence of the plasmid
XX pRx-Bcl-xl-bsr, which contains the human Bcl-xl gene, and is used in an
XX example from the present invention
XX
XX Sequence 7372 BP; 2353 A; 1749 C; 1649 G; 1621 T; 0 U; 0 Other;
XX
XX Query Match 7.9%; Score 76.4; DB 2; Length 7372;
XX Best Local Similarity 45.2%; Pred. No. 0.00012;
XX Matches 281; Conservative 0; Mismatches 341; Indels 0; Gaps 0;
XX
XX 11 AGAAATAGCAGCAATATTTTATAAAAGCATGCAATTTCTTTATAGATCGGAAGTTTAA 70
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 4550 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4609
XX
XX 71 AAAAAACATAGAAATGTTCAATATTACATGGGTTTTTATGATACATGCAATAT 130
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 4610 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4669
XX
XX 131 TTATTTATTTTCATGAGTTTTTTATGGATAGCATGACAAATATTATATCATGTTAAT 190
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 4670 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4729
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 191 AACATGTTTGTCTTAAATACATGCAATTTTAAATCAGACATTTCTTTAAATCAAA 250
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
Db 4730 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4789
Qy 251 TCTAATCTCTTATACACACGACATTTCAGCGAAAAATTCAGTAAAAAGAGAAAAATAAG 310
Db 4790 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4849
Qy 311 AATGAGAGATAGAGAGATTTCTATGGAAGAAAGAGAGAGAACATGTAGGTGAACAAAT 370
Db 4850 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4909
Qy 371 AAAGAGATATGATGATATATTTTATGAGAGGTGGTGAAGATATTTTATGAGAGGAGAG 430
Db 4910 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4969
Qy 431 AGAATAGAAAAAGAAAAATGACATGGTGAATCTGAAGAAGATGAATTTGTGTTAAAGATGA 490
Db 4970 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 5029
Qy 491 AGAGAGAAAGAGAACTCCATGGCTTAAAGTCTCGTAAAGAGATGAAGAAACAAAGA 550
Db 5030 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 5089
Qy 551 AGGAAGAGAAAGAGAAAGGCTAAATAGACTAACTATTGCCAAAAATTTCTGTAGCCGAC 610
Db 5090 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 5149
Qy 611 AAATACTATTGTTGGTCCAAAGTTT 632
Db 5150 AAAAAAGATTGTCGAGCTT 5171

RESULT 15
AA333180
ID AA333180 standard; DNA; 7797 BP.
XX
AC AA333180;
XX
DT 25-JUN-1999 (first entry)
DE Cowpox virus bsr full length gene sequence.
XX
KW Cowpox virus; bsr; viral vector; expression; apoptosis; resistance; crmA;
KW bcl-2; bcl-xl; FLIP; survivin; IAP; ILP; adenovirus; cancer;
KW autoimmune disease; graft rejection reaction; inflammation;
XX inflammatory disease; ss.
XX
OS Cowpox virus.
XX
PN WO9913073-A2.
XX
PD 18-MAR-1999.
XX
PF 07-SEP-1998; 98WO-JP004010.
XX
PR 08-SEP-1997; 97JP-00259235.
XX
PA (RPRG-) RPR GENCELL ASIA PACIFIC INC.
XX
PI Hamada H;
XX
DR WPI; 1999-243728/20.
XX
PT New apoptosis-resistant virus-sensitive cell.
XX
PS Example 1; Page 34-38; 51pp; English.
XX
CC The present invention describes an apoptosis-resistant virus-sensitive
CC cell line into which an apoptosis resistance gene has been introduced.
CC The recombinant viruses generated are capable of expressing apoptosis-
CC associated genes. These can then be used in a variety of diseases for
CC which the induction of apoptosis by gene transfer, or where the
CC inhibition of harmful apoptosis, is therapeutic. The recombinant viruses
CC are useful as vectors for gene therapy which can be applied to cancer
```

CC therapy for destroying cancer cells selectively, the treatment of
CC autoimmune diseases and graft rejection reaction, and apoptosis induction
CC therapy for inflammatory cells in inflammatory diseases. Prior arts have
CC encountered the problem where if an adenovirus vector capable of
CC expressing an apoptosis-associated gene is introduced into animal cells,
CC the cells producing the virus will be destroyed because the period of
CC time required to induce cell death by apoptosis is shorter than that
CC required to replicate and produce the virus, resulting in failure to
CC obtain a recombinant virus having the integrated apoptosis-associated
CC gene. In this invention an apoptosis-resistant 293 cell line (having an
CC apoptosis resistant gene introduced) is established and overcomes the
CC problem. The present sequence represents the cowpox virus bsr gene which
CC is used in an example from the present invention
XX
SQ

Sequence 7797 BP; 2542 A; 1760 C; 1656 G; 1839 T; 0 U; 0 Other;

Query Match	7.9%	Score 76.4;	DB 2;	Length 7797;
Best Local Similarity	45.2%;	Pred. No. 0.00012;		
Matches 281;	Conservative 0;	Mismatches 341;	Indels 0;	Gaps 0;
Qy 11	AGAAATGACGAGAAATTTATTA	AAAGCATGCAATCTCTTATAGATCGGAAGTTTAA	70	
Db 4975	AAAAAAAAAAAAAAAAAAAAA	AAAAAAAAAAAAAAAAAAAAA	5034	
Qy 71	AAAAACATATAGATTGTTACA	TATTTACATGGGTTTTTATTGGATAACATGACAAATAT	130	
Db 5035	AAAAAAAAAAAAAAAAAAAAA	AAAAAAAAAAAAAAAAAAAAA	5094	
Qy 131	TTATTTATTTTCATGAGTTT	TATTTGGATGATGACAAATATTATATATATCATGTGTTAAT	190	
Db 5095	AAAAAAAAAAAAAAAAAAAAA	AAAAAAAAAAAAAAAAAAAAA	5154	
Qy 191	AACATGTTTGGTCTTAAATA	TACATGCTTTTAAATCAGACATTTGTTTTTAAATCAAA	250	
Db 5155	AAAAAAAAAAAAAAAAAAAAA	AAAAAAAAAAAAAAAAAAAAA	5214	
Qy 251	TCTAATCTCTTATATACAC	GACATTTGACGGAAATTCAGGTAAGAGAAATAAAG	310	
Db 5215	AAAAAAAAAAAAAAAAAAAAA	AAAAAAAAAAAAAAAAAAAAA	5274	
Qy 311	AATGAGAGATAGAGAGATTT	CTATGGAAAAAGAGAGACATGTAGGTGAACAAAT	370	
Db 5275	AAAAAAAAAAAAAAAAAAAAA	AAAAAAAAAAAAAAAAAAAAA	5334	
Qy 371	AAAGAGATATGATGATATTT	TATGAGAGGTGGTGAAGATTATTTTAGAGAGGGAGAG	430	
Db 5335	AAAAAAAAAAAAAAAAAAAAA	AAAAAAAAAAAAAAAAAAAAA	5394	
Qy 431	AGAAATAGAAAAAGAAATG	ACATGCTGAAGAATGAAATTTGTTTAAAGATGA	490	
Db 5395	AAAAAAAAAAAAAAAAAAAAA	AAAAAAAAAAAAAAAAAAAAA	5454	
Qy 491	AGAGAGAAAGAGAACTCC	ATGGCTAAAGTCTCGTAAAGAGATGAAAAAGAAACAAAGA	550	
Db 5455	AAAAAAAAAAAAAAAAAAAAA	AAAAAAAAAAAAAAAAAAAAA	5514	
Qy 551	AGGAAGAAAGAAAGAGCT	AAATATAGACTAATTCGCCAAATTTCTGTAGCCGAC	610	
Db 5515	AAAAAAAAAAAAAAAAAAAAA	AAAAAAAAAAAAAAAAAAAAA	5574	
Qy 611	AAATACTATTTGGTCCA	GGTT 632		
Db 5575	AAAAAGAAATTTGGTCA	AGCTT 5596		

Search completed: November 7, 2005, 15:14:47
Job time : 653 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2005, 07:25:35 ; Search time 3913 Seconds
(without alignments)
9357.993 Million cell updates/sec

Title: US-10-800-161-24

Perfect score: 962

Sequence: 1 tgtgtttctcagaataagca.....aacattttcttctgctgaac 962

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 segs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1:.*
2: gb_est2:.*
3: gb_hc:.*
4: gb_est3:.*
5: gb_est4:.*
6: gb_est5:.*
7: gb_est6:.*
8: gb_gss1:.*
9: gb_gss2:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	231	24.0	267	8	BH854168 SALK_0788
2	222	23.1	228	8	BZ352867
3	152	15.8	152	8	BZ770600 SALK_1435
4	149.6	15.6	222	9	CC797914 SALK_1455
5	147	15.3	327	8	CC455857
6	144.6	15.0	203	8	BZ381166
7	128.4	13.3	387	8	BH749275 SALK_0475
8	117.4	12.2	223	8	BZ359042 SALK_1337
9	110.6	11.5	237	8	BH814587 SALK_0666
10	101.2	10.5	147	8	BZ352865
11	99.8	10.4	1101	9	CNS0021J
12	99.6	10.4	1434	8	BZ576813
13	98.2	10.2	860	7	CF289324
14	97.2	10.1	1201	9	CNS0167M
15	96	10.0	660	8	BH183498
16	96	10.0	660	9	CNS070NJ
17	95.4	9.9	1267	9	AG346514
18	95	9.9	1235	9	AG361148
19	94	9.8	737	9	CL077131
20	93.2	9.7	764	7	CF289381
21	92.6	9.6	982	8	AQ325799
22	92.6	9.6	1228	9	CU104752
23	92.4	9.6	1061	9	CU145135
24	92.4	9.6	1241	9	AG448181

25	92.4	9.6	1599	9	CL083840
26	92.4	9.6	1872	8	BZ558518
27	92	9.6	1101	9	CNS0039G
28	91.8	9.5	1047	8	CC266341
29	91.2	9.5	842	9	CNS010QN
30	91.2	9.5	1409	9	AG346237
31	91	9.5	826	7	CV483860
32	90.8	9.4	795	7	CF288665
33	90.4	9.4	1036	9	CNS00599
34	90.4	9.4	1355	9	AG346348
35	90.2	9.4	934	8	AZ184244
36	89.8	9.3	1025	9	CL113815
37	89.6	9.3	997	9	CL112076
38	89.6	9.3	1059	9	CNS0022B
39	89.6	9.3	1110	9	CL037168
40	89.4	9.3	755	7	CV491618
41	89.4	9.3	1074	8	BZ696936
42	89.4	9.3	1210	9	CG749728
43	89.4	9.3	1251	9	AG332167
44	89.4	9.3	1522	9	CL128484
45	89.2	9.3	1614	9	AG365768

ALIGNMENTS

RESULT 1
LOCUS BH854168 267 bp DNA linear GSS 13-JUN-2002
DEFINITION SALK_078835.55.00.x Arabidopsis thaliana TDNA insertion lines
Arabidopsis thaliana genomic clone SALK_078835.55.00.x, genomic survey sequence.

ACCESSION BH854168
VERSION BH854168.1 GI:21425039

KEYWORDS GSS.

SOURCE Arabidopsis thaliana (thale cress)

ORGANISM Arabidopsis thaliana

REFERENCE 1 (bases 1 to 267)
Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R.,
Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L.,
Shinn,P., Zimmerman,J. and Ecker,J.R.

AUTHORS

TITLE

JOURNAL

COMMENT

Arabidopsis Genome

Unpublished (2001)

Contact: Joseph R. Ecker

Salk Institute Genomic Analysis Laboratory (SIGnAL)

The Salk Institute for Biological Studies

10010 N. Torrey Pines Road, La Jolla, CA 92037, USA

Tel: 858 453 4100 x1752

Fax: 858 558 6379

Email: ecker@salk.edu

This is single pass sequence recovered from the left border of TDNA.

Class: TDNA tagged.

Location/Qualifiers

source

1. .267

/organism="Arabidopsis thaliana"

/mol_type="genomic DNA"

/ecotype="Col-0"

/db_xref="taxon:3702"

/clone="SALK_078835.55.00.x"

/notes="PCR was performed on Arabidopsis thaliana lines

each of which contains one or more TDNA insertion

elements. The resultant fragment for each line was

directly sequenced to determine the genomic sequence at

the site of insertion. Details of the protocols used can

be found at http://signal.salk.edu/tdna_protocols.html"

ORIGIN

Query Match 24.0%; Score 231; DB 8; Length 267;
 Best Local Similarity 94.0%; Pred. No. 2.6e-32;
 Matches 251; Conservative 0; Mismatches 15; Indels 1; Gaps 1;

QY 23 AATATTTTATAAAAGCATCAATTCCTTTATAGATCGCGAAGTTTAA-AAAAACATATA 81
 |||||
 DB 1 ATATATTTTATAAAGCATCAATTCCTTTATACATCGCGAAGTTTAAAAACATATA 60
 |||||

QY 82 GAATTGTTTACAATATTACATGGGTTTTTTATTGGATAACATGACAAATATTATTATTTC 141
 |||||
 DB 61 GAATTGTTACAATATTACATGAGTTTTTTATTGGATAACATGACAAATATTATATATTAC 120
 |||||

QY 142 ATGAGTTTTTTTGGATAGCATGACAAATATTATATATATATCAGTGTTAATAACATGTTTTG 201
 |||||
 DB 121 ATGAATTTTTTGTGGATAGCATGACAAATATTATATATATCAGTGTTAATAACATGTTTTG 180
 |||||

QY 202 TTTCTTAAATACATGCATTTTAAATCAGACATTTGTTTTTAAATCAAAATCTAATCTCTT 261
 |||||
 DB 181 TTTCTTAAATACATGCATTTTAAATCAGACATTTGTTTTTAAATCAAAATATAATCTCTT 240
 |||||

QY 262 ATATCACACGACATTCGACGAAAATT 288
 |||||
 DB 241 ATATCACACGACATTCGATGAAGATT 267
 |||||

RESULT 2
 BZ352867
 LOCUS
 DEFINITION
 SALK_082724.45.05.x Arabidopsis thaliana genomic clone SALK_082724.45.05.x, genomic survey sequence.

ACCESSION
 VERSION BZ352867.1 GI:24943729
 KEYWORDS
 SOURCE
 ORGANISM
 Arabidopsis thaliana (thale cress)
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 228)
 Alonzo,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R.,
 Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L.,
 Shinn,P., Zimmerman,J. and Ecker,J.R.
 A Sequence-Indexed Library of Insertion Mutations in the
 Arabidopsis Genome
 Unpublished (2001)
 Contact: Joseph R. Ecker
 Salk Institute Genomic Analysis Laboratory (SIGnAL)
 The Salk Institute for Biological Studies
 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
 Tel: 858 453 4100 x1752
 Fax: 858 558 6379
 Email: ecker@salk.edu
 This is single pass sequence recovered from the left border of
 TDNA. This sequence lies within 300 bases of the 5' end of
 At3g25882.
 Class: TDNA tagged.

FEATURES
 source
 Location/Qualifiers
 1..228
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /ecotype="Col-0"
 /db_xref="taxon:3702"
 /clone="SALK_082724.45.05.x"
 /clone_lib="Arabidopsis thaliana TDNA insertion lines"
 /note="PCR was performed on Arabidopsis thaliana lines
 each of which contains one or more TDNA insertion
 elements. The resultant fragment for each line was
 directly sequenced to determine the genomic sequence at
 the site of insertion. Details of the protocols used can
 be found at http://signal.salk.edu/tdna_protocols.html

ORIGIN
 Query Match 15.8%; Score 152; DB 8; Length 152;
 Best Local Similarity 100.0%; Pred. No. 7.6e-18;
 Matches 152; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match 23.1%; Score 222; DB 8; Length 228;
 Best Local Similarity 97.4%; Pred. No. 1.2e-30;
 Matches 222; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 701 TTTTTCACACATGGACTTCCTTTATTTCAAAAAGTCAATAAAGTGTGACGTGATGATACT 760
 |||||
 DB 1 TTTTTCNCNCNTGGACTTCCTTTATTCNAAAGTCNATANAGTGTGACGTGATGATACT 60
 |||||

QY 761 TAGCCTTTAAACATCGCATGATGATGTCTATTAGCATCAATCTCCACCGTCCCAATTTATT 820
 |||||
 DB 61 TAGCCTTTAAACATCGCATGATGATGTCTATTAGCATCAATCTCCACCGTCCCAATTTATT 120
 |||||

QY 821 TAGTTGTTGACAATATCGACCGTCTAAGTTTCCACCGCGGTATTAAGAGTTTCATTAT 880
 |||||
 DB 121 TAGTTGTTGACAATATCGACCGTCTAAGTTTCCACCGCGGTATTAAGAGTTTCATTAT 180
 |||||

QY 881 AATATTTTACCAAAATAAATCAGCAATAATTTTTTCTTGACTAAGCT 928
 |||||
 DB 181 AATATTTTACCAAAATAAATCAGCAATAATTTTTTCTTGACTAAGCT 228
 |||||

RESULT 3
 BZ770600
 LOCUS
 DEFINITION
 SALK_143535.56.00.x Arabidopsis thaliana TDNA insertion lines
 Arabidopsis thaliana genomic clone SALK_143535.56.00.x, genomic survey sequence.

ACCESSION
 VERSION BZ770600.1 GI:28944284
 KEYWORDS
 SOURCE
 ORGANISM
 Arabidopsis thaliana (thale cress)
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 152)
 Alonzo,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R.,
 Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L.,
 Shinn,P., Zimmerman,J. and Ecker,J.R.
 A Sequence-Indexed Library of Insertion Mutations in the
 Arabidopsis Genome
 Unpublished (2001)
 Contact: Joseph R. Ecker
 Salk Institute Genomic Analysis Laboratory (SIGnAL)
 The Salk Institute for Biological Studies
 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
 Tel: 858 453 4100 x1752
 Fax: 858 558 6379
 Email: ecker@salk.edu
 This is single pass sequence recovered from the left border of
 TDNA. This sequence lies within 300 bases of the 5' end of
 At3g25882.
 Class: TDNA tagged.

FEATURES
 source
 Location/Qualifiers
 1..152
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /ecotype="Col-0"
 /db_xref="taxon:3702"
 /clone="SALK_143535.56.00.x"
 /clone_lib="Arabidopsis thaliana TDNA insertion lines"
 /note="PCR was performed on Arabidopsis thaliana lines
 each of which contains one or more TDNA insertion
 elements. The resultant fragment for each line was
 directly sequenced to determine the genomic sequence at
 the site of insertion. Details of the protocols used can
 be found at http://signal.salk.edu/tdna_protocols.html

ORIGIN
 Query Match 15.8%; Score 152; DB 8; Length 152;
 Best Local Similarity 100.0%; Pred. No. 7.6e-18;
 Matches 152; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 777 GCATGATGTCATTAGCATCAATCTCCACCGTCCAAATTTATTTAGTTGTTGACAATAT 836
Db 1 GCATGATGTCATTAGCATCAATCTCCACCGTCCAAATTTATTTAGTTGTTGACAATAT 60

QY 837 CGACCGTCTAAGTCTCCACACCGCTATAGAGTTTCATTTATAAATTTTAGCAAAATA 896
Db 61 CGACCGTCTAAGTCTCCACACCGCTATAGAGTTTCATTTATAAATTTTAGCAAAATA 120

QY 897 AAATCAGCAAAATAATTTTCTTGACTAAGCT 928
Db 121 AAATCAGCAAAATAATTTTCTTGACTAAGCT 152

RESULT 4
LOCUS CC797914.1 222 bp DNA linear GSS 01-JUL-2003
DEFINITION SALK_145557.51.25.x Arabidopsis thaliana TDNA insertion lines
Arabidopsis thaliana genomic clone SALK_145557.51.25.x, genomic
survey sequence.
ACCESSION CC797914.1 GI:32393137
VERSION CC797914
KEYWORDS Arabidopsis thaliana (thale cress)
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (bases 1 to 222)
REFERENCE Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R.,
AUTHORS Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L.,
Shinn,P., Zimmerman,J. and Ecker,J.R.
TITLE A Sequence-Indexed Library of Insertion Mutations in the
JOURNAL Arabidopsis Genome
COMMENT Unpublished (2001)
Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGnAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of
TDNA.
Class: TDNA tagged.
FEATURES Location/Qualifiers
source 1..222
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/ecotype="Col-0"
/db_xref="taxon:3702"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/note="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tdna_protocols.html"

ORIGIN
Query Match 15.6%; Score 149.6; DB 9; Length 222;
Best Local Similarity 85.3%; Pred. No. 2.1e-17;
Matches 191; Conservative 0; Mismatches 26; Indels 7; Gaps 2;

QY 285 AATTCAGGTAAAGAGAAATAAAGAAATGAGATGAGAGATTTCTATGGAAGAAAGAA 344
Db 222 AATTCAGGTAAAGAGAAATAAAGAAATGAGATGAGAGATTTCTATGGAAGAAAGAA 165

QY 345 AGAGAGAACATGTAGGTGACAAATAAAGAGATATGATGATATTTTATGAGAGGTGG 404
Db 164 AGAGGGAACATGTGGGTGGACAAATAAAGAGATATGATGATATTTTATGAGAGGTGG 105

QY 405 TGAAGATTATTT-----TAGGAGAGGAGAGAGAAATAGAAAAAGAAAAATGACATGCTGA 459

```

```

Db 104 TGAAGATTATTTAGGAGAGGAGAGAGAAATAGAAATAGAAAAAGAAATGACATGCTGA 45
QY 460 ATCTGAAGAAGATGAATTTGTTTAAAGATGAAGAGAGAGAAAGAGA 503
Db 44 ATCTGAAGAAGATGAATTTGTTTAAAGAGAGAGAGAGAAAGAGA 1

RESULT 5
LOCUS CC455857 327 bp DNA linear GSS 30-MAY-2003
DEFINITION SALK_089493.54.85.x Arabidopsis thaliana TDNA insertion lines
Arabidopsis thaliana genomic clone SALK_089493.54.85.x, genomic
survey sequence.
ACCESSION CC455857
VERSION CC455857.1 GI:31216228
KEYWORDS Arabidopsis thaliana (thale cress)
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (bases 1 to 327)
REFERENCE Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R.,
AUTHORS Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L.,
Shinn,P., Zimmerman,J. and Ecker,J.R.
TITLE A Sequence-Indexed Library of Insertion Mutations in the
JOURNAL Arabidopsis Genome
COMMENT Unpublished (2001)
Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGnAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of
TDNA.
Class: TDNA tagged.
FEATURES Location/Qualifiers
source 1..327
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/ecotype="Col-0"
/db_xref="taxon:3702"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/note="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tdna_protocols.html"

ORIGIN
Query Match 15.3%; Score 147; DB 8; Length 327;
Best Local Similarity 93.7%; Pred. No. 6.1e-17;
Matches 164; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

QY 1 TGTGTTTCTCAGAAATAGCAGAAATATTTATAAAAAAGCATGCAATTTCTTTATAGATCG 60
Db 153 TGTGTTTCTCAGAAATAGCAGATATATTTATAAAAAAGCATGCAATTTCTTTATAAATCG 212

QY 61 CGAAGT-TTAAAAAACATATAGAAATTTGTACAAATTTACATGCGGTTTTTTATGGAATAC 119
Db 213 CGAAGTCTTTAAGAACAATAGAAATTTGTACAAATTTACATCGGTTTTTTATGGAATAC 272

QY 120 ATGACAAATATTTATTTATTTTCATGAGTTTTTTATGATAGCATGACAAATATTTA 174
Db 273 ATGACTAATATTTATATATTTTCATGAGTTTTTTGTTGATAGCATGACAAATATTTA 327

RESULT 6

```

```

ACCESSION      BH749275
VERSION        BH749275.1  GI:18963963
SOURCE         GSS.
ORGANISM       Arabidopsis thaliana (thale cress)
                Arabidopsis thaliana
                Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
                rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE      1 (bases 1 to 387)
AUTHORS        Alonso,J.M., Letesche,T.J., Barajas,P., Chen,H., Cheuk,R.,
                Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L.,
                Shinn,P., Zimmerman,J. and Ecker,J.R.
TITLE          A Sequence-Indexed Library of Insertion Mutations in the
                Arabidopsis Genome
JOURNAL        Unpublished (2001)
COMMENT        Contact: Joseph R. Ecker
                Salk Institute Genomic Analysis Laboratory (SIGnAL)
                The Salk Institute for Biological Studies
                10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
                Tel: 858 453 4100 x1752
                Fax: 858 558 6379
                Email: ecker@salk.edu
                This is single pass sequence recovered from the left border of
                TDNA.
FEATURES       TDNA tagged.
                Location/Qualifiers
                1..387
                /organism="Arabidopsis thaliana"
                /mol_type="genomic DNA"
                /ecotype="Col-0"

```

```

/db_xref="taxon:3702"
/clone="SAUK_047543_32_95.x"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/note="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tdna\_protocols.html"

Query Match      13.3%;   Score 128.4;   DB 8;   Length 387;
Best Local Similarity 89.8%;   Pred. No. 1.5e-13;
Matches 149; Conservative 0; Mismatches 16; Indels 1; Gaps 1

ORIGIN

```

Qy	285	AATTCAGGTTAAAGAGAAATTAAGATGAGAGATAGAGAGATTTCTATGGAAGAA	344
Db	344	AATTCAGGTTAAAGAGAAATTAAGATGAGAGATAGAGATTTCTATGGAATGAA	285
Qy	345	AGAGAGAACATCTAGGTGAAC-AAAATAAAGAGATATGATGATATATTTTATGAGGGTG	403
Db	284	AGAGAGAACATCTGGGTGATCAAAAATAAAGAGATATGATGATATATTTTATGAGGGTG	225
Qy	404	GTCAAGATTATTTTAGGACGGGAGAGAGAAATAGAAAAAGAAAAT	449
Db	224	GTCAAGATTATCATAGGACGGTCGGTGACATGAATCAGCAATT	179
RESULT 8			
BZ359042/c			
LOCUS	BZ359042	223 bp	DNA linear
DEFINITION	SALK_133702.17.85.n Arabidopsis thaliana tDNA insertion lines		GSS 14-NOV-200
	Arabidopsis thaliana genomic clone SALK_133702.17.85.n, genomic		
	survey sequence.		
ACCESSION	BZ359042		

SOURCE	ORGANISM	REFERENCE
Arabidopsis thaliana (thale cress)		
Arabidopsis thaliana		
Eukaryota; Viridiplantae		
Streptophyta; Embryophyta; Tracheophyta		
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosoids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.		
1 (bases 1 to 223)		

AUTHORS

Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R.,
Gadrinab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L.,
Shinn, P., Zimmerman, J. and Ecker, J.R.
A Sequence-Indexed Library of Insertion Mutations in the
Arabidopsis Genome
Unpublished (2001)

TITLE

JOURNAL
COMMENT

Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGNAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu

This is single pass sequence recovered from the left border of
TDNA.

Class: TDNA tagged.

FEATURES
source

Location/Qualifiers
1..223

/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/ecotype="Col-0"
/db_xref="taxon:3702"

/clone="SALK_133702.17.85.n"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/note="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tdna_protocols.html"

ORIGIN

Query Match 12.2%; Score 117.4; DB 8; Length 223;

Best Local Similarity 76.7%; Pred. No. 1.6e-11;

Matches 171; Conservative 0; Mismatches 46; Indels 6; Gaps 2;

Qy 285 AATTCAGTAAAGAGAGAAAATAAAGATGAGATAGAGAGATTCTATGGAAGAA 344

Db 223 AATTCAGCAAAACGGAAGCAATGCATGCGACATAGAGCTATTTCTATCCAAATAC 164

Qy 345 AGACAGAA-CATGTAGTGAACAAATAAAGAGATATGATATATATTTATGAGGTG 403

Db 163 AGAGTGGATCATGTGGTTCAACAACATACAGAGATATGATCTATATTTCTGAGAGATG 104

Qy 404 GTGAAGATTATTT-----TAGGAGGGGAGAGAGAAATAGAAAAGAAATGCATGGT 458

Db 103 GGACAGATATTTCTTTGAGGGGAGAGAGATAGAAAATAGAAAAGAAATGCATGGT 44

Qy 459 AATCTGAAGAGATCAATTGTTTAAAGATGAAGAGAGAAAGA 501

Db 43 AGTCTCAAGATACATCGTTTTCAGACCGCAGAGAGAAAGA 1

RESULT 9

BH814587

LOCUS

DEFINITION BH814587 237 bp DNA linear GSS 02-MAY-2002
SALK_066674 Arabidopsis thaliana TDNA insertion lines Arabidopsis
thaliana genomic clone SALK_066674, genomic survey sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Arabidopsis thaliana

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis

1 (bases 1 to 237)

Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R.,

Gadrinab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L.,

Shinn, P., Zimmerman, J. and Ecker, J.R.

A Sequence-Indexed Library of Insertion Mutations in the

Arabidopsis Genome

Unpublished (2001)

COMMENT

Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGNAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu

This is single pass sequence recovered from the left border of
TDNA.

Class: TDNA tagged.

FEATURES
source

Location/Qualifiers
1..237

/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/ecotype="Col-0"
/db_xref="taxon:3702"

/clone="SALK_066674"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/note="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tdna_protocols.html"

ORIGIN

Query Match 11.5%; Score 110.6; DB 8; Length 237;

Best Local Similarity 92.8%; Pred. No. 2.8e-10;

Matches 116; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 833 ATATCGACCGCTAAGTTCCACCGCGGTATTAAGAGTTTCTATTATAAATTTTAGCAA 892

Db 1 ATATCGACCGCTAAGTTCCACCGCGGTATTAAGAGTTTCTATTATAAATTTTAGCAA 60

Qy 893 AATAAAATCAGCAAAATAATTTTTTCTGACTTAAGCTTAAACGAGCGCCGTTAACATTTTCT 952

Db 61 AATAAAATCAGCAAAATAATTTTTTCTGACTTAAGCTTAAACGAGCGCGGTTAACATTTCT 120

Qy 953 TCTGG 957

Db 121 TGTGG 125

RESULT 10

BZ352865

LOCUS

DEFINITION BZ352865 147 bp DNA linear GSS 14-NOV-2002
SALK_082722.32.45.x Arabidopsis thaliana TDNA insertion lines
Arabidopsis thaliana genomic clone SALK_082722.32.45.x, genomic
survey sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Arabidopsis thaliana

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis

1 (bases 1 to 147)

Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R.,

Gadrinab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L.,

Shinn, P., Zimmerman, J. and Ecker, J.R.

A Sequence-Indexed Library of Insertion Mutations in the

Arabidopsis Genome

Unpublished (2001)

Contact: Joseph R. Ecker

Salk Institute Genomic Analysis Laboratory (SIGNAL)

The Salk Institute for Biological Studies

10010 N. Torrey Pines Road, La Jolla, CA 92037, USA

Tel: 858 453 4100 x1752

Fax: 858 558 6379

Email: ecker@salk.edu

This is single pass sequence recovered from the left border of

TDNA. This sequence lies within 300 bases of the 5' end of

JOURNAL
COMMENT

J. Bacteriol. (2002) In press
Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 20622216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.

FEATURES
source

Location/Qualifiers
1..1434
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="MSH"
/db_xref="taxon:287"
/clone="msh2_5115"
/clone_lib="msh"
/notes="Environmental isolate. Whole genomic shotgun library."

ORIGIN

Query Match 10.4%; Score 99.6; DB 8; Length 1434;
Best Local Similarity 39.4%; Pred. No. 2.7e-08;
Matches 270; Conservative 0; Mismatches 414; Indels 2; Gaps 1;

QY 11 AGAAATAGCAGCAATATTTATATAAAGCATGCAATTCCTCTTAGATGCGAAGTTTAA 70
DB 1378 ATAAATAATATATATATATTAATAATAATAATAATAATAATAATAATAATAATAA 1319
QY 71 AAAACATATAGAAATGTTCAATATATACATGCGGTTTTTATATGATAA--CATGACAAAT 128
DB 1318 AAAAATAATAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 1259
QY 129 ATTATTTATTCATGAGTTTTTTCGATAGCATGACAAATATATATATCATGCTTA 188
DB 1258 ATTATATATATTAATAAATAATATATAAATTTAAATAATAATAATAATAATAA 1199
QY 189 ATAACATCTTTTGTCTTAAAATACATGCAATTTTAAAATCAGACATTTGTTTAAATCA 248
DB 1198 ATTATATTAATATATATAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1139
QY 249 AATCTAATCTCTATATACACGACATTCGACGAAATTCAGGTAAAAAGAGAAATAA 308
DB 1138 ANAAAAAANANTATAAANAATAAANAATAAANAATAAANAATAAANAATAAANA 1079
QY 309 AGAATGAGATAGAGAGATTTCTATCGAATAAAGAAAGAGAGACATGTAGTGACAAA 368
DB 1078 NNAANAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1019
QY 369 ATAAAGAGATATGATGATATATTTTATGAGAGGTGGTGAAGATTATTTTAGGAGGGAG 428
DB 1018 AAANAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 959
QY 429 AGAGAAATAGAAAAGAAATATGATCGTGAATCTGGAAGAGATGAATTTGTTAAAGAT 488
DB 958 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 899
QY 489 GAAGAGAGAAGAGAACTCCATGCTAAAGTCTGTTAAGAAGATGAAAAGAAACAAA 548
DB 898 AANAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 839
QY 549 GAAGGAAGAAGAAGAGAAAGGCTTAAATATAGACTATTTGCCAAATTTCTGTAGCCG 608
DB 838 AANAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 779
QY 609 ACAAACTACTATTTGGTCCAGGTTATTTTGTGATTTCTTTTGAAGTCAAAAGTTATTTCT 668
DB 778 ANNNNANNNANNN 719
QY 669 TACATATCTCTAAATAATAGCCGA 694
DB 718 NNN 693

RESULT 13
CF289324LOCUS
DEFINITION

CF289324 860 bp mRNA linear EST 14-AUG-2003
IMAGE:4971130 5', mRNA sequence.

ACCESSION
VERSION

CF289324.1 GI:33649103

KEYWORDS
SOURCE

EST.
Xenopus laevis (African clawed frog)

ORGANISM

Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Xenopus; Xenopus.

REFERENCE

1 (bases 1 to 860)

AUTHORS

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL

Unpublished (1997)

COMMENT

Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgaps-k@mail.nih.gov
Tissue Procurement: Dr. Igor Dawid
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Cloned through the I.M.A.G.E. Consortium
www-bio.llnl.gov/bbrp/image/image.html
Plate: LLAM10955 row: p column: 11
High quality sequence start: 333
High quality sequence stop: 571.
Location/Qualifiers
1..860
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="IMAGE:4971130"
/dev_stage="embryo, stage 31-32"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NICHD XGC Emb4"
/notes="Organ: whole embryo; Vector: pCMV-SPORT6; Site: 1;
NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo
dT. Average insert size 2.1 kb. Constructed by Life
Technologies. Note: This is a Xenopus Gene Collection
(XGC) library."

FEATURES
source

ORIGIN

Query Match 10.2%; Score 98.2; DB 7; Length 860;
Best Local Similarity 49.7%; Pred. No. 4.9e-08;
Matches 250; Conservative 0; Mismatches 253; Indels 0; Gaps 0;

QY 94 TATTACATGGTGTTCATGATAACATGACAAATATTTATTTTATTTTATTTTATTTAT 153
DB 70 TATTTTCTTTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 129
QY 154 TGGATGACATGACAAATATTAATATATCATGTGTTAATAACATGTTTGTCTTAAATAC 213
DB 130 TTTATATATAAAAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAA 189
QY 214 ATGCATTTTAAATCAGACATTTGTTTAAATCAATCTTAATCTTATATCACACGA 273
DB 190 AAAAAATTTTATTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 249
QY 274 CATTTGCGGAAAAATTCAGGTAAAAAGAGAAAAATAAAGATGAGAGATGAGAGATTTCTA 333
DB 250 AATATATAAATAAATTAATTTAAAAAATAAATAAATAAATAAATAAATAAATAA 309
QY 334 TGAAGAAAGAGAGAGACATGTAGGTGAACAAAAATAAGAGATATGATGATATATTTT 393
DB 310 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 369

```
Qy 394 ATGAGAGGTGGTGAAGATTATTTTAGAGAGGAGAGAGAGAAATAGAAAAAGAAAAATGACA 453
Db 370 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 429
Qy 454 TGGTGAATCTGAAGAGATCAATTTGTTTAAAGATGAAGAGAGAGAAAGAGAACTCCATGCC 513
Db 430 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 489
Qy 514 TAAAGGCTCTGTAAGAGATGAAAAAGAAAAAGAAAGAGAGAGAAAGAGAAAGGCTA 573
Db 490 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 549
Qy 574 AATAGACTTAATTTGTCCTCAAAA 596
Db 550 AAAAAAAAAAAAAAAAAACCAACAAA 572

RESULT 14
CNS0167M/c
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence T7 end of BAC
BACN15M24 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
VERSION ALI06396.1 GI:5621701
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 1201)
Genoscope.
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : sefre@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelOBAC11.
FEATURES
source Location/Qualifiers
1..1201
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN15M24"
/clone_lib="DrosBAC"
/plasmid="pBelOBAC11"
/note="end : T7"

ORIGIN
Query Match 10.1%; Score 97.2; DB 9; Length 1201;
Best Local Similarity 36.8%; Pred. No. 7.4e-08;
Matches 228; Conservative 84; Mismatches 307; Indels 0; Gaps 0;

Qy 11 AGAAATAGCAGGAATATTTATAAAAAGCATGCAATTCCTTATAGATCGCGAAGTTAA 70
Db 1155 ATAWNAANAANTATATAAAAAAATAWATATATAAAWAAAAAAWATATATTAATAWAA 1096
Qy 71 AAAAACATATAGATTGTTCAATATTACATGGGTTTTTATTGATCAACATGACAAATAT 130
Db 1095 TATAANWAAAAAATAAAAAAANWTTTTHANAAATATTTTWTNATATAWATTTT 1036
Qy 131 TTAATTTTTCATGAGTTTTTATTGGATGACATGACAAATATTAATATATCAGTGTAA 190
Db 1035 TTTTATTTTATATAWAAAAWAAAAAATTTTAAAAAATAAAWTAATTTATWAAAAATTT 976
```

```
Qy 191 AACATGTTTTTCTTAAATATACATGCATTTTAAATACAGACATTTTGTTTTAAAAATCAAA 250
Db 975 TAAAAATTTTWTATWTTTTTAAAAAATAATATWAAAWTTTTTTTATATWTTATAA 916
Qy 251 TCTAATCTCTTATATACAAACGACATTGACGGAAAAATTCAGGTAAAAAGAGAAAAATAAAG 310
Db 915 WAWAWTTTTTTTWTAAAAAATAATTTAAAAATWTAAAAATTTATAAATAAAAAA 856
Qy 311 AATGAGAGATAGAGAGATTTCTATGAAAAAGAAAGAGAGACATCTAGGTGTAACAAAT 370
Db 855 AAAAAAATAAAAAAATAATTTTAAATATATAAATTTTAAAAAATAAAAAAATAAATTT 796
Qy 371 AAAGAGATATGATATATTTTATCAGAGGTGCTGAAGATTTATTTTAGAGAGGAGAG 430
Db 795 TTTTAAAAATAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 736
Qy 431 AGAAATAGAAAAAGAAATGCATGCTGTAATCTGAAGAAGATGAATTTGCTTTAAAGATGA 490
Db 735 AAAAAWATTTWATATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAA 676
Qy 491 AGAGAGAAAGAGAACTCCATGCTGCTAAAGTCTCTGAAGAAGATGAAAAAGAAACAAAGA 550
Db 675 MAMMYMMCAAAAMMAAAVMAAAAMACMAWMTAAAGCNTAAAAAACAACAAAMAAAMAM 616
Qy 551 AGGAAGAAGAAAGAAAGGCTAAATAGACTAACTATTGCCCCAAAATTTCTGTAGCGAC 610
Db 615 AAATAAAMAAATCTAKKNDKKKWAADAKDDNAKASAWGBKKWMAATWACATGGCGCAB 556
Qy 611 AATACTATTTTGGTCCAAAG 629
Db 555 AMAWTTGWATTGCAAAWS 537

RESULT 15
BH183498/c
LOCUS
DEFINITION
023_L07-rev SmbAC1 Schistosoma mansoni genomic clone 023L07 5',
genomic survey sequence.
ACCESSION
VERSION BH183498.1 GI:16288814
KEYWORDS GSS.
SOURCE Schistosoma mansoni
ORGANISM Schistosoma mansoni
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
Strigeidida; Schistosomatidae; Schistosomatidae; Schistosoma.
REFERENCE
1 (bases 1 to 660)
Le Paslier,M.C., Pierce,R.J., Merlin,F., Hirai,H., Wu,W.,
Williams,D.L., Johnston,D., LoVerde,P.T. and Le Paslier,D.
Construction and characterization of a Schistosoma mansoni
bacterial artificial chromosome library
Genomics 65 (2), 87-94 (2000)
10783255
Other GSSs: 023 L.07-21
Contact: Pierce RJ
INSERM U 167
Institut Pasteur de Lille
1 rue du Professeur A. Calmette, 59019-Lille, France
Tel: (33) (0)3 20877783
Fax: (33) (0)3 20877888
Email: Raymond.Pierce@pasteur-lille.fr
CNS sequencing ID=DG0AA023CF04BP1
Plate: 023 row: L column: 07
Seq primer: M13 reverse primer
Class: BAC ends
High quality sequence stop: 660.
FEATURES
source Location/Qualifiers
1..660
/organism="Schistosoma mansoni"
/mol_type="genomic DNA"
/strain="Puerto-Rican"
/db_xref="taxon:6183"
/clone="023L07"
```


This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2005, 05:19:40 ; Search time 976 Seconds
(without alignments)
8151.195 Million cell updates/sec

Title: US-10-800-161-24
Perfect score: 962
Sequence: 1 tgtgtttctcagaatagca.....aacattttcttctggttaac 962

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 9794790 seqs, 4134909567 residues

Total number of hits satisfying chosen parameters: 19589580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
18: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
19: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq.*
20: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq.*
21: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq.*
22: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq.*
23: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
24: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq2.*
25: /cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq.*
26: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
27: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
28: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	962	100.0	962	20	US-10-800-161-24
2	862	89.6	862	20	US-10-800-161-25
3	862	89.6	1700	20	US-10-800-161-3
4	862	89.6	1700	21	US-10-760-752-3
5	544	56.5	544	20	US-10-800-161-27
					Sequence 25, Appl
					Sequence 2, Appl
					Sequence 3, Appl
					Sequence 3, Appl
					Sequence 27, Appl

6	274	28.5	274	20	US-10-800-161-26	Sequence 26, Appl
7	272.4	28.3	274	20	US-10-800-161-28	Sequence 82, Appl
8	86	8.9	537	20	US-10-021-323-8120	Sequence 8120, Ap
9	81	8.4	1214	19	US-10-424-599-102083	Sequence 102083,
10	80.8	8.4	3673778	17	US-10-312-841-2	Sequence 2, Appli
11	79.6	8.3	2512	13	US-09-925-065A-27191	Sequence 27191, A
12	79.6	8.3	2512	13	US-09-925-065A-27192	Sequence 27192, A
13	79.2	8.2	2512	13	US-09-925-065A-27190	Sequence 27190, A
14	77.8	8.1	921	21	US-10-425-115-38710	Sequence 38710, A
15	76.8	8.0	7657	15	US-10-239-676-185	Sequence 185, App
16	76.8	8.0	7657	16	US-10-311-455-1995	Sequence 1995, Ap
17	76.8	8.0	14006	16	US-10-311-455-1931	Sequence 1931, Ap
18	76.8	8.0	1223197	14	US-10-027-632-179264	Sequence 179264,
19	76.6	8.0	1223197	18	US-10-027-632-179264	Sequence 179264,
20	76.6	8.0	529	20	US-10-021-323-10810	Sequence 10810, A
21	74.8	7.8	9539	15	US-10-239-676-52	Sequence 52, Appl
22	74.8	7.8	9539	16	US-10-240-453-54	Sequence 54, Appl
23	74.4	7.7	627	20	US-10-021-323-9336	Sequence 9336, Ap
24	74.4	7.7	2520	18	US-10-389-566-205	Sequence 205, App
25	74.2	7.7	517	20	US-10-021-323-11054	Sequence 11054, A
26	73.6	7.7	529	9	US-09-983-965-2109	Sequence 2109, Ap
27	73.6	7.7	1297	21	US-10-425-115-67687	Sequence 67687, A
28	73.4	7.6	113515	16	US-10-311-455-2147	Sequence 2147, Ap
29	73.2	7.6	478	20	US-10-021-323-6774	Sequence 6774, Ap
30	73	7.6	1029	21	US-10-425-115-123086	Sequence 123086,
31	72.8	7.6	1204	20	US-10-437-963-77858	Sequence 77858, A
32	72.6	7.5	16033	16	US-10-311-455-1377	Sequence 1377, Ap
33	72.4	7.5	803	20	US-10-437-963-72176	Sequence 72176, A
34	72.4	7.5	3673778	17	US-10-312-841-1	Sequence 1, Appli
35	72.2	7.5	778	21	US-10-363-348A-2179	Sequence 2179, Ap
36	72.2	7.5	778	21	US-10-363-348A-2180	Sequence 2180, Ap
37	72.2	7.5	778	22	US-10-363-483A-2179	Sequence 2179, Ap
38	72.2	7.5	778	22	US-10-363-483A-2180	Sequence 2180, Ap
39	72	7.5	446	9	US-09-960-352-3400	Sequence 3400, Ap
40	72	7.5	16167	16	US-10-311-455-1056	Sequence 1056, Ap
41	72	7.5	16167	16	US-10-240-485-82	Sequence 82, Appl
42	71.6	7.4	996	21	US-10-425-115-151933	Sequence 151933,
43	71.6	7.4	17934	16	US-10-311-455-1692	Sequence 1692, Ap
44	71.6	7.4	462586	24	US-10-476-264-420	Sequence 420, App
45	71.4	7.4	1759	13	US-09-925-065A-47393	Sequence 47393, A

ALIGNMENTS

RESULT 1
US-10-800-161-24
; Sequence 24, Application US/10800161
; Publication No. US20040154051A1
; GENERAL INFORMATION:
; APPLICANT: Cade, Rebecca M
; APPLICANT: Dietrich, Robert A
; APPLICANT: Lawton, Kay Ann
; TITLE OF INVENTION: INDUCIBLE PROMOTERS
; FILE REFERENCE: A-31089CIP1
; CURRENT APPLICATION NUMBER: US/10/800,161
; CURRENT FILING DATE: 2004-03-12
; PRIOR APPLICATION NUMBER: 60/171,008
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: 60/175,519
; PRIOR FILING DATE: 2000-01-11
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
; LENGTH: 962
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-10-800-161-24

Query Match 100.0%; Score 962; DB 20; Length 962;
Best Local Similarity 100.0%; Pred. No. 2.7e-167;
Matches 962; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	TGTGTTTCTCAGAAATAGCAGAAATATTTATAAAAAAGCATGCAATTCCTCTATTAGATCG	60
Db	1	TGTGTTTCTCAGAAATAGCAGAAATATTTATAAAAAAGCATGCAATTCCTCTATTAGATCG	60
Qy	61	CGAAGTTTAAAAAACAATATAGAAATGGTTACAAATATACATGGTGTTTTATTGGATAACA	120
Db	61	CGAAGTTTAAAAAACAATATAGAAATGGTTACAAATATTTACATGGTGTTTTATTGGATAACA	120
Qy	121	TGCAAAATATTTATTTATTTTCATGATGTTTATTTGATAGCATGCAAAATATTAATATAT	180
Db	121	TGCAAAATATTTATTTATTTTCATGAGTTTATTTGATAGCATGCAAAATATTAATATAT	180
Qy	181	CAGTGTAAATAAATGATGTTTGTCTTAAATACATGCAATTTTAAATCAGCAATTTGTTT	240
Db	181	CAGTGTAAATAAATGATGTTTGTCTTAAATACATGCAATTTTAAATCAGCAATTTGTTT	240
Qy	241	TAAATCAAATCTAATCTCTTATATACAAACACATTTGACGAAATTCAGGTGAAAAAGA	300
Db	241	TAAATCAAATCTAATCTCTTATATCAACACATTTGACGAAATTCAGGTGAAAAAGA	300
Qy	301	GAAATATAAGAAATGAGAGATAGAGAGATTTCTATGAAAAAGAAAGAGAGACATGTAGG	360
Db	301	GAAATATAAGAAATGAGAGATAGAGAGATTTCTATGAAAAAGAAAGAGAGACATGTAGG	360
Qy	361	TGAACAAATATAAGAGATATGATGATATATTTTATGAGAGGTGGTGAAGATTTATTTTAGG	420
Db	361	TGAACAAATATAAGAGATATGATGATATATTTTATGAGAGGTGGTGAAGATTTATTTTAGG	420
Qy	421	AGGGGAGAGAAATAGAAAAAGAAATGACATGCTGAATCTGAAGAAGATGAAATGTG	480
Db	421	AGGGGAGAGAAATAGAAAAAGAAATGACATGCTGAATCTGAAGAAGATGAAATGTG	480
Qy	481	TTAAGATGAAAGAGAGAAAGAGAACTCCATGCTAAAGTCTCGTAAAGAAGATGAAAAAG	540
Db	481	TTAAGATGAAAGAGAGAAAGAGAACTCCATGCTAAAGTCTCGTAAAGAAGATGAAAAAG	540
Qy	541	AAACAAAAGAGGAAGAGAAAGAGAAAGGCTAAAAATAGACTAACTATTGGCCAAAATTTTC	600
Db	541	AAACAAAAGAGGAAGAGAAAGAGAAAGGCTAAAAATAGACTAACTATTGGCCAAAATTTTC	600
Qy	601	TGTAGCCGACAAATACTATTTGGTCCAAAGTTATTTTGTGTATCTCTTTTGAAGTCAAAAG	660
Db	601	TGTAGCCGACAAATACTATTTGGTCCAAAGTTATTTTGTGTATCTCTTTTGAAGTCAAAAG	660
Qy	661	TTATTTCTTACATATCTCTAAAAATATAGCCGATACCAATTTTCCACACATGGACTTC	720
Db	661	TTATTTCTTACATATCTCTAAAAATATAGCCGATACCAATTTTCCACACATGGACTTC	720
Qy	721	CTTTATTTCCAAAGTCAATAAAGTGTGACGTCATGATCTTACGCTTTTAAACATCGCAT	780
Db	721	CTTTATTTCCAAAGTCAATAAAGTGTGACGTCATGATCTTACGCTTTTAAACATCGCAT	780
Qy	781	GATGATGTCATTAGCATCAATCTCCACCGTCCAAATTTATTAGTTGTTGACAAATATCGAC	840
Db	781	GATGATGTCATTAGCATCAATCTCCACCGTCCAAATTTATTAGTTGTTGACAAATATCGAC	840
Qy	841	CGTCTAAGTTCCACACCGACGGCTATAGAGTTTCATTTAAATTTTACGAAAAATAAAT	900
Db	841	CGTCTAAGTTCCACACCGACGGCTATAGAGTTTCATTTAAATTTTACGAAAAATAAAT	900
Qy	901	CAGCAAAATAATTTTTTCTTGACTAAGCTTTAAACGACGCCGCTTAAACATTTTCTTGCTTA	960
Db	901	CAGCAAAATAATTTTTTCTTGACTAAGCTTTAAACGACGCCGCTTAAACATTTTCTTGCTTA	960
Qy	961	AC 962	
Db	961	AC 962	

RESULT 2

US-10-800-161-25

00-10-000-101-23 ; Sequence 25, Application US/10800161

; Publication No. US20040154051A1


```
Db 241 TTGAAGTCAAAAGTTATTTCTTACATATATCTCTAAAAATATAGCCGATACCAATTTTTC 300
Qy 708 ACATCGACTTCCTTTATTTCCAAAGTCAATAAGTGTGACGTCATGATCTTACGCTT 767
Db 301 ACATCGACTTCCTTTATTTCCAAAGTCAATAAGTGTGACGTCATGATCTTACGCTT 360
Qy 768 TAAACATCGCATGATGATGTCATTAGCATCAATCTCCACCGTCCAAATTTTATTTAGTTGT 827
Db 361 TAAACATCGCATGATGATGTCATTAGCATCAATCTCCACCGTCCAAATTTTATTTAGTTGT 420
Qy 828 TGCAATATCGACCGTCTAGTTCCACACGACCGCTTAAAGAGTTTCATTAATAATTTT 887
Db 421 TGCAATATCGACCGTCTAGTTCCACACGACCGCTTAAAGAGTTTCATTAATAATTTT 480
Qy 888 AGCAAAATAAAATCAGCAATAATTTTCTTGACTAAGCTTAAACGACCGGTTAACAT 947
Db 481 AGCAAAATAAAATCAGCAATAATTTTCTTGACTAAGCTTAAACGACCGGTTAACAT 540
Qy 948 TTTC 951
Db 541 TTTC 544
```

```
RESULT 6
US-10-800-161-26
; Sequence 26, Application US/10800161
; Publication No. US20040154051A1
; GENERAL INFORMATION:
; APPLICANT: Cade, Rebecca M
; APPLICANT: Dietrich, Robert A
; APPLICANT: Lawton, Kay Ann
; TITLE OF INVENTION: INDUCIBLE PROMOTERS
; FILE REFERENCE: A-31089CIP1
; CURRENT APPLICATION NUMBER: US/10/800,161
; CURRENT FILING DATE: 2004-03-12
; PRIOR APPLICATION NUMBER: 60/171,008
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: 60/175,519
; PRIOR FILING DATE: 2000-01-11
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 274
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-10-800-161-26
```

```
Query Match 28.5%; Score 274; DB 20; Length 274;
Best Local Similarity 100.0%; Pred. No. 7.5e-41;
Matches 274; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 678 TCTAAAAATATAGCCGATACCAATTTTCCACACATGCGACTTCCTTTATTTCCAAAAGTCA 737
Db 1 TCTAAAAATATAGCCGATACCAATTTTCCACACATGCGACTTCCTTTATTTCCAAAAGTCA 60
Qy 738 ATAAAGTGTGACGTCATGATCTTACGCTTTTAAACATCGCATGATCTGCTATTAGCAT 797
Db 61 ATAAAGTGTGACGTCATGATCTTACGCTTTTAAACATCGCATGATCTGCTATTAGCAT 120
Qy 798 CAATCTCCACCGTCCAAATTTTATTTAGTTGTGACAATATCGACGCTTAAAGTTCCACACC 857
Db 121 CAATCTCCACCGTCCAAATTTTATTTAGTTGTGACAATATCGACGCTTAAAGTTCCACACC 180
Qy 858 GACGGCTATAAGAGTTTCATTATAAATTTTAGCAAAATAAAATCAGCAATAAATTTTTC 917
Db 181 GACGGCTATAAGAGTTTCATTATAAATTTTAGCAAAATAAAATCAGCAATAAATTTTTC 240
Qy 918 TTGACTAAGCTTAAACGACCGGTTAACATTTTC 951
Db 241 TTGACTAAGCTTAAACGACCGGTTAACATTTTC 274
```

```
RESULT 7
US-10-800-161-28
; Sequence 28, Application US/10800161
; Publication No. US20040154051A1
; GENERAL INFORMATION:
; APPLICANT: Cade, Rebecca M
; APPLICANT: Dietrich, Robert A
; APPLICANT: Lawton, Kay Ann
; TITLE OF INVENTION: INDUCIBLE PROMOTERS
; FILE REFERENCE: A-31089CIP1
; CURRENT APPLICATION NUMBER: US/10/800,161
; CURRENT FILING DATE: 2004-03-12
; PRIOR APPLICATION NUMBER: 60/171,008
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: 60/175,519
; PRIOR FILING DATE: 2000-01-11
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
; LENGTH: 274
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-10-800-161-28
```

```
Query Match 28.3%; Score 272.4; DB 20; Length 274;
Best Local Similarity 99.6%; Pred. No. 1.5e-40;
Matches 273; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 678 TCTAAAAATATAGCCGATACCAATTTTCCACACATGCGACTTCCTTTATTTCCAAAAGTCA 737
Db 1 TCTAGAAATATAGCCGATACCAATTTTCCACACATGCGACTTCCTTTATTTCCAAAAGTCA 60
Qy 738 ATAAAGTGTGACGTCATGATCTTACGCTTTTAAACATCGCATGATCTGCTATTAGCAT 797
Db 61 ATAAAGTGTGACGTCATGATCTTACGCTTTTAAACATCGCATGATCTGCTATTAGCAT 120
Qy 798 CAATCTCCACCGTCCAAATTTTATTTAGTTGTGACAATATCGACGCTTAAAGTTCCACACC 857
Db 121 CAATCTCCACCGTCCAAATTTTATTTAGTTGTGACAATATCGACGCTTAAAGTTCCACACC 180
Qy 858 GACGGCTATAAGAGTTTCATTATAAATTTTAGCAAAATAAAATCAGCAATAAATTTTTC 917
Db 181 GACGGCTATAAGAGTTTCATTATAAATTTTAGCAAAATAAAATCAGCAATAAATTTTTC 240
Qy 918 TTGACTAAGCTTAAACGACCGGTTAACATTTTC 951
Db 241 TTGACTAAGCTTAAACGACCGGTTAACATTTTC 274
```

```
RESULT 8
US-10-021-323-8120
; Sequence 8120, Application US/10021323
; Publication No. US20040123340A1
; GENERAL INFORMATION:
; APPLICANT: Deikman, Jill
; APPLICANT: Feng, Paul C.C.
; APPLICANT: Fincher, Karen L.
; APPLICANT: Ziegler, Todd E.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(52274)B
; CURRENT APPLICATION NUMBER: US/10/021,323
; CURRENT FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: US 60/255, 619
; PRIOR FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 17880
; SEQ ID NO 8120
; LENGTH: 537
; TYPE: DNA
; ORGANISM: Gossypium hirsutum
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3828-006-Q1-N6-D9
US-10-021-323-8120
```



```
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27190
; LENGTH: 2512
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-27190

Query Match      8.2%; Score 79.2; DB 13; Length 2512;
Best Local Similarity 54.5%; Pred. No. 0.00011;
Matches 156; Conservative 1; Mismatches 129; Indels 0; Gaps 0;

QY 283 AAAATTTCAGGTAAAGAGAGAAATAAAGATGAGAGATAGAGAGATTCTTATCGAAAAAG 342
DB 690 AAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAGAGAGAGAGAGAGAGAGAA 749
QY 343 AAAGAGAGAACATGTAGGTGAAACAAATAAAGAGATATGATATATATTTATGAGAGGT 402
DB 750 AGAGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 809
QY 403 GGTGAAGATTATTTAGGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATC 462
DB 810 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 869
QY 463 TGAAGAAGATGAATTGTGTTAAAGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 522
DB 870 AAAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 929
QY 523 GTAAAGAGAGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 568
DB 930 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 975

RESULT 14
US-10-425-115-38710/c
; Sequence 38710, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 38710
; LENGTH: 921
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(921)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_135307C.1
US-10-425-115-38710

Query Match      8.1%; Score 77.8; DB 21; Length 921;
Best Local Similarity 45.7%; Pred. No. 0.00014;
Matches 271; Conservative 0; Mismatches 322; Indels 0; Gaps 0;
```

```
QY 23 AAATATTTATAAAAGCATGCAATTCTCTTATAGATCGCGAAGTTTAAAAAACAATATAG 82
DB 861 AAAAAATATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAA 802
QY 83 AATTGTTACAATATTACATGGGTTTTTTATTTGGATACATGACAAATATTATTATTTC 142
DB 801 AAAAAAAATAATAAAAAAAATAAAAAAAATAAAAAAAATAAAAAAAATAAAAAAA 742
QY 143 TGAGTTTTTATTTGGATAGCATGACAATATTATATATATCAGTGTATTAATCACTGTTTGT 202
DB 741 AAAAAAAAGAGAAAAAAAGAAAAAGAAAAAAAGAAAAAAAGAAAAAAAGAAAAAA 682
QY 203 TCTTAAAAATACATGCATTTTAAATCAGACATTTGTTTAAATCAAAATCTAATCTCTTA 262
DB 681 AAAAAAAAGAAAAAAAGAAAAAAAGTAAAAAAAGTAAAAAAAGTAAAAAAAGTAAAA 622
QY 263 TATCACACGACATTCACGGAATAATTCAGGTAAAAAGAGAAAAATAAGATGAGAGATAG 322
DB 621 AAAAAATAAAAAAAGAAAAAAAGAAAAAAAGAAAAAAAGAAAAAAAGAAAAAAAG 562
QY 323 AGAGATTTCTATGAAAAAGAGAGAGAGAACATGTAGGTCAACAAAAATAAGAGATATGA 382
DB 561 CAAAAAAACAAAAAAAGAGAAAAAAAGAAAAAAAGAAAAAAAGAAAAAAAGAAAAAA 502
QY 383 TGATATATTTTATGAGAGGTGTTGAAGATTTATTTAGGAGAGGAGAGAGAGAGAGAG 442
DB 501 AAAAAAAAGAAAAATACAAAAAAAGAAAAAAAGAAAAAAAGAAAAAAAGAAAAAAAG 442
QY 443 AGAAATGACATGCTGAATCTGAGAGAGATGAATTTGTTTAAAGATGAGAGAGAGAGAG 502
DB 441 AAAAAAGAAAAAGTAGAATAATAAAAAAAAGAAAAAAAGAAAAAAAGAAAAAAAGAAA 382
QY 503 AACTCATGCTAAAGTCTCGTAAAGAGATGAAAAAGAAAAACAAAAAGAGAGAGAGAAA 562
DB 381 AACATAAAATAAAATACAAATAAAAAAAAGAAAAAAAGAAAAAAAGAAAAAAAGTTGACAA 322
QY 563 GAGAAAGGCTAAATAAGACTAACTATTGCCAAAAATTTCTGTAGCCCAAAATA 615
DB 321 CAAAAAAAGAAAAAGAAAAAAAGAAAAATTAACAAAAATTTATTCGTAAGAAAAA 269
```

```
RESULT 15
US-10-239-676-185/c
; Sequence 185, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
; FILE REFERENCE: 5013.1003
; CURRENT APPLICATION NUMBER: US/10/239,676
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: PCT/EP01/03968
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-04-06
; 2000-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 185
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-185
```


This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2005, 02:50:19 ; Search time 217 Seconds
(without alignments)
7253.909 Million cell updates/sec

Title: US-10-800-161-24
Perfect score: 962
Sequence: 1 tgtgtttctcagaatagca.....aacattttcttctggctaac 962

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 81813859 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	862	89.6	1700	US-09-733-685-3	Sequence 3, Appli
2	103.2	10.7	7218	US-08-232-463-14	Sequence 14, Appl
3	76	7.9	612	US-09-902-540-1357	Sequence 1357, Ap
4	76	7.9	19124	US-08-487-8268-13	Sequence 13, Appl
5	72.2	7.5	12703	US-09-949-016-16685	Sequence 16685, A
6	71.2	7.4	152331	US-09-128-155-16	Sequence 16, Appl
7	71.2	7.4	176373	US-09-128-155-17	Sequence 17, Appl
8	69	7.2	133559	US-09-949-016-15845	Sequence 15845, A
9	67.2	7.0	36731	US-09-949-016-13770	Sequence 13770, A
10	67.2	7.0	168174	US-10-071-411A-63	Sequence 63, Appl
11	67.2	7.0	168273	US-10-071-411A-2	Sequence 2, Appli
12	66	6.9	96922	US-09-949-016-17061	Sequence 17061, A
13	66	6.9	288775	US-09-949-016-16435	Sequence 16435, A
14	65.8	6.8	147382	US-09-949-016-14624	Sequence 14624, A
15	64.8	6.7	55841	US-09-949-016-16602	Sequence 16602, A
16	64.8	6.7	248968	US-09-949-016-12614	Sequence 12614, A
17	64.8	6.7	250958	US-09-949-016-16061	Sequence 16061, A
18	64.4	6.7	601	US-09-949-016-30531	Sequence 30531, A
19	64.4	6.7	601	US-09-949-016-37150	Sequence 37150, A
20	64.4	6.7	601	US-09-949-016-37154	Sequence 37154, A
21	64.4	6.7	601	US-09-949-016-145868	Sequence 145868, A
22	64.4	6.7	601	US-09-949-016-146136	Sequence 146136, A
23	64.4	6.7	601	US-09-949-016-146404	Sequence 146404, A
24	64.4	6.7	205044	US-09-949-016-15851	Sequence 15851, A
25	64.4	6.7	205044	US-09-949-016-15852	Sequence 15852, A
26	64.4	6.7	205044	US-09-949-016-15853	Sequence 15853, A
27	64.4	6.7	223471	US-09-949-016-12387	Sequence 12387, A

28	64.4	6.7	223471	4	US-09-949-016-12724	Sequence 12724, A
29	64.4	6.7	223471	4	US-09-949-016-12725	Sequence 12725, A
30	63.8	6.6	509	4	US-09-733-685-1	Sequence 1, Appli
31	62.6	6.5	601	4	US-09-949-016-184864	Sequence 184864, A
32	62.6	6.5	601	4	US-09-949-016-184865	Sequence 184865, A
33	62.6	6.5	601	4	US-09-949-016-184866	Sequence 184866, A
34	62.6	6.5	601	4	US-09-949-016-184867	Sequence 184867, A
35	62.6	6.5	601	4	US-09-949-016-184868	Sequence 184868, A
36	62.6	6.5	601	4	US-09-949-016-184869	Sequence 184869, A
37	62.6	6.5	601	4	US-09-949-016-184870	Sequence 184870, A
38	62.6	6.5	601	4	US-09-949-016-184871	Sequence 184871, A
39	62.6	6.5	601	4	US-09-949-016-185019	Sequence 185019, A
40	62.6	6.5	601	4	US-09-949-016-185020	Sequence 185020, A
41	62.6	6.5	601	4	US-09-949-016-185021	Sequence 185021, A
42	62.6	6.5	601	4	US-09-949-016-185022	Sequence 185022, A
43	62.6	6.5	601	4	US-09-949-016-185023	Sequence 185023, A
44	62.6	6.5	601	4	US-09-949-016-185024	Sequence 185024, A
45	62.6	6.5	601	4	US-09-949-016-185025	Sequence 185025, A

ALIGNMENTS

RESULT 1
US-09-733-685-3
; Sequence 3, Application US/097333685
; Patent No. 6706952
; GENERAL INFORMATION:
; APPLICANT: Cade, Rebecca M
; APPLICANT: Dietrich, Robert A
; TITLE OF INVENTION: GENES ENCODING PROTEINS INVOLVED IN THE REGULATION OF
; FILE REFERENCE: A-31089A
; CURRENT APPLICATION NUMBER: US/09/733,685
; CURRENT FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/171,008
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: 60/175,519
; PRIOR FILING DATE: 2000-01-11
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1700
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (365)..(374)
; OTHER INFORMATION: TCAl motif
; NAME/KEY: misc feature
; LOCATION: (426)..(435)
; OTHER INFORMATION: TCAl motif
; NAME/KEY: misc feature
; LOCATION: (609)..(614)
; OTHER INFORMATION: MYCATR22 element
; NAME/KEY: misc feature
; LOCATION: (646)..(665)
; OTHER INFORMATION: CAMV ASI salicylic acid response element
; NAME/KEY: misc feature
; LOCATION: (707)..(712)
; OTHER INFORMATION: PAL BOX
; NAME/KEY: misc feature
; LOCATION: (757)..(762)
; OTHER INFORMATION: HEXAMERAT 4 element
; NAME/KEY: misc feature
; LOCATION: (863)..(1228)
; OTHER INFORMATION: Nil6 genomic coding region
US-09-733-685-3

Query Match 89.6%; Score 862; DB 4; Length 1700;
Best Local Similarity 100.0%; Pred. No. 5e-175;
Matches 862; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 38766 ATATGTATTATCTATTAAATATATATGTATTTATCTATTAAATATATATATATATATAT 38707
Qy 164 GACAAATATTAAATATATCAGTGTGTTAAATAACATGTTTGTCTTAAATATACATGCTTTTA 223
Db 38706 -AAATATATATATAATATATAAATAATATATTTATATATTAATATATAAATATATTTT 38648
Qy 224 AAATCAGACATTTGTTTAAATCAATCTTAATCTCTATATCAACG-ACATGTACGG 282
Db 38647 ATATATTAAATATATAACATATATTTATATATCAAAATATATAAATATATAAATATAT 38588
Qy 283 AAAATTCCAGGTAAAAAGAGAAATAAAGAAATGAGAGATAGAGATCTTCTATGAAAAAG 342
Db 38587 ATCAATATATATAACATATATTTATATATCAAAATATATAAATATATAAATATATCAAA 38528
Qy 343 AAAGAGAGAACATGTAGGTGCAACAAAATAAAGAGATATGATGATATATTTATGAGAGGT 402
Db 38527 TATATATAACATATTTATATATCAATATATATAAATATATTTATATATCAATATATA 38468
Qy 403 GGTGAAGATTATTTTAGGAGGGGAGAGAGAAATAGAAAAAGAAATGACATGGTGAATC 462
Db 38467 TAAAAACATATTTATATATCAATATATATAAATATATTTATATATCAATATATAATA 38408
Qy 463 TGAAGAAGATCAATGTGTTAAAGATCAAGAGAGAGAAAGAACTCCATGCTAAAGTCTC 522
Db 38407 AAATATATTTATATATTTAAATATATATATAAATATATATAAATATATATAATGTATG 38348
Qy 523 GTAAAGAGATGAAAAAGAAACAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 580
Db 38347 GCAAAACAGATAAATAAATCTGAAAGGCAAAACAGATAAATCCAAATAAACAGTGA 38290

RESULT 13

US-09-949-016-16435
; Sequence 16435, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16435
; LENGTH: 258775
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16435

Query Match 6.9%; Score 66; DB 4; Length 258775;
Best Local Similarity 51.7%; Pred. No. 0.00016;
Matches 150; Conservative 0; Mismatches 140; Indels 0; Gaps 0;
Qy 295 AAAAGAGAAATAAAGATGAGAGATAGAGAGATTTCTATGAAAAAGAAAGAGAGAA 354
Db 122511 AAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 122570
Qy 355 TGTAGGTGAACAAATTAAGAGATATGATATATTTTATGAGAGGTGGTGAAGATTAT 414
Db 122571 AGGAGAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 122630
Qy 415 TTTAGGAGCGGAGAGAGAAATAGAAAAAGAAATGACATGGTCAATCTCTGAAGAGATGA 474
Db 122631 GAGAGAGAAAGAAAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 122690
Qy 475 ATTGTGTTAAAGATGAAGAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 534

Db 122691 AGAGGAGAGAGAAAAAGAGAGAAAAAGAGAGAAAAAGAGAGAAAAAGAGAGAA 122750
Qy 535 AAAAGAGAAACAAAAG 584
Db 122751 AGAAGAGAAAGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 122800

RESULT 14

US-09-949-016-14624/c
; Sequence 14624, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14624
; LENGTH: 147382
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc feature
; LOCATION: (1)-(147382)
; OTHER INFORMATION: n = A, T, C or G
US-09-949-016-14624

Query Match 6.8%; Score 65.8; DB 4; Length 147382;
Best Local Similarity 47.1%; Pred. No. 0.00016;
Matches 202; Conservative 0; Mismatches 227; Indels 0; Gaps 0;
Qy 11 AGAAATAGCAGCAAAATATTTTATAAAGCATGCAATCTCTTATAGATCCGGAAGTTTAA 70
Db 137087 AAATATATATAAT 137028
Qy 71 AAAACATATAGAAATTTGTACATATATACATGCGTGTGTTTATTCGATACATGACAAATAT 130
Db 137027 TAAATATATAATATATAAATATATATATATATATATATATATATATATATATAT 136968
Qy 131 TTATTTTATTTTCATGAGTGTGTTTATTTGGATAGCATGACAAATATTAATATATATCATGTGTTTAT 190
Db 136967 ATATAAATATGTTAAATATATACATATATATAAATATATATAAATATATATAAATAT 136908
Qy 191 AACATGTTTGTCTTTAAATATACATGCAATTTTAAAAATCGACATCTTTGTTTAAAAATCAAA 250
Db 136907 ATATAACTAATAATATAAATATATATAAATATATAAATATATAAATATATATATATAT 136848
Qy 251 TCTAATCTCTTATATACACACGACATTCACGGAATTCAGGTAAAAGAGAGAAATTAAG 310
Db 136847 TAAATATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATAT 136788
Qy 311 AATGAGAGATAGAGAGATTTCTATGAAAAAGAAAGAGAGAAACATGTAGGTGAACAAAT 370
Db 136787 ATATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATATATAAAT 136728
Qy 371 AAAGAGATATGATATATATTTTATGAGAGGTGGTGAAGATTATTTTATGAGAGGGAGAG 430
Db 136727 AAATATATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATAT 136668
Qy 431 AGAAATAGA 439
Db 136667 ATATATATA 136659

RESULT 15

US-09-949-016-16602/c
; Sequence 16602, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16602
; LENGTH: 55841
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-16602

Query Match 6.7%; Score 64.8; DB 4; Length 55841;
Best Local Similarity 53.8%; Pred. No. 0.00021;
Matches 155; Conservative 0; Mismatches 132; Indels 1; Gaps 1;

QY 294 AAAAAGAGAAATATAAGATGAGAGATAGAGAGATTTCTATGGAAAAAGAAAGAGAGAAC 353
DB 31510 AAAGAGAGAGAGAAAGGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAG 31451

QY 354 ATGTAGGTGAACAAATATAAGAGATATGATGATATATTTTATGAGAGGTGGTGAAGATTA 413
DB 31450 AAAAGAAAGAAAGAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 31391

QY 414 TTTTAGGAGAGGAGAGAGAAATAGAAAAAGAAAAATGACATGGTGAATCTTGAAGAGATG 473
DB 31390 AAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 31331

QY 474 AATTGTGTTAAAGATGAGAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAT 533
DB 31330 GAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 31271

QY 534 GAAAAAGAAACAAAAAGAGGAG-AGAAAGAGAGAAAGGCTAAAAATAGA 580
DB 31270 AAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 31223

Search completed: November 7, 2005, 06:51:43
Job time : 219 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2005, 06:51:50 ; Search time 4632 Seconds
(without alignments)
10063.454 Million cell updates/sec

Title: US-10-800-161-24
Perfect score: 962
Sequence: 1 tgggtttctcagaatagca.....aacattttcttctggctaac 962

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

1: gb_ba.*
2: gb_hgt.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	962	100.0	83650	8 AB023041	Arabidopsis
C 2	862	89.6	1700	6 AR488147	Sequence
C 3	676.2	70.3	92620	8 AB026636	Arabidopsis
C 4	632.2	65.7	83646	8 AB005248	Arabidopsis
C 5	593.8	61.7	94487	8 AC012394	Arabidopsis
C 6	593.8	61.7	100806	8 AC015450	Arabidopsis
C 7	579.2	60.2	104386	8 ATT32A17	Arabidopsis
C 8	579.2	60.2	179771	8 ATCHRIV25	Arabidopsis
C 9	571.6	59.4	95519	8 AF071527	Arabidopsis
C 10	571.6	59.4	116448	8 AC005142	Arabidopsis
C 11	571.6	59.4	159629	8 ATCHRIV9	Arabidopsis
C 12	406.8	42.3	95190	8 AC007203	Arabidopsis
C 13	120.6	12.5	105223	8 AC007399	Arabidopsis
C 14	103.2	10.7	7218	6 I66494	Sequence 14
C 15	99.2	10.3	19577	10 BX255966	Mouse DNA
C 16	94.2	9.8	1197	8 AJ840663	Arabidopsis
C 17	92.6	9.6	175366	10 AC101718	Arabidopsis
C 18	90.2	9.4	173049	10 AC115878	Mus muscu
C 19	90	9.4	286208	2 AC117140	Rattus no

C 20	89.2	9.3	191209	10 AC125184	Mus muscu
C 21	88.8	9.2	171258	2 AC116819	Mus muscu
C 22	87.6	9.1	164443	2 CR751221	Danio rer
C 23	87.4	9.1	161115	2 AC101939	Mus muscu
C 24	87.4	9.1	174536	10 AC125079	Mus muscu
C 25	87.4	9.1	216464	10 AL844515	Mouse DNA
C 26	86.8	9.0	191001	2 AC138586	Mus muscu
C 27	86.8	9.0	202872	2 AC016160	Homo sapi
C 28	86.6	9.0	165720	10 AC127244	Mus muscu
C 29	86.4	9.0	110000	8 CR382131.03	Continuation (4 of
C 30	86.2	9.0	218313	10 AC132465	Mus muscu
C 31	86	8.9	175787	10 AC142103	Mus muscu
C 32	85.6	8.9	43736	9 AC130469	Homo sapi
C 33	85.6	8.9	44735	9 AC092315	Homo sapi
C 34	85.6	8.9	64789	2 AC083839	Homo sapi
C 35	85.6	8.9	227625	2 AC094806	Rattus no
C 36	85.6	8.9	239189	2 AC097180	Rattus no
C 37	85	8.8	178038	10 AL672308	Mouse DNA
C 38	84.6	8.8	422	5 PD0M04	X93505 P.domesticu
C 39	84.6	8.8	164094	2 AC101966	Mus muscu
C 40	84.6	8.8	176666	2 AC132131	Mus muscu
C 41	84.6	8.8	177649	10 AL732328	Mouse DNA
C 42	84.4	8.8	159453	10 AC126558	Mus muscu
C 43	84.4	8.8	224248	2 AC124316	Mus muscu
C 44	84.2	8.8	245524	2 AC095456	Rattus no
C 45	84.2	8.8	270712	2 AC137360	Rattus no

ALIGNMENTS

RESULT 1
AB023041/c
LOCUS AB023041 83650 bp DNA linear PLN 14-FEB-2004
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 3, PI clone: MPE11.
ACCESSION AB023041 BA000014
VERSION AB023041.1 GI:4220640
KEYWORDS Arabidopsis thaliana (thale cress)
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1
AUTHORS Sato, S., Nakamura, Y., Kaneko, T., Katoh, T., Asamizu, E. and Tabata, S.
TITLE Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence features of the regions of 4,504,864 bp covered by sixty PI and TAC clones
JOURNAL DNA Res. 7 (2), 131-135 (2000)
MEDLINE 20277480
PUBMED 10819329
REFERENCE 2 (bases 1 to 83650)
AUTHORS Sato, S., Nakamura, Y., Kaneko, T., Katoh, T., Asamizu, E. and Tabata, S.
TITLE Direct Submission
JOURNAL Submitted (01-FEB-1999) Yasukazu Nakamura, Kazusa DNA Research Institute, Department of Plant Gene Research, 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: ynakamu@kazusa.or.jp, Tel: 81-438-52-3935, Fax: 81-438-52-3934)
COMMENT Address for correspondence: kaos@kazusa.or.jp
For the latest information on annotation of this clone, please see http://www.kazusa.or.jp/kaos/cgi-bin/sgd/graph.cgi?c=MPE11
Genes with similarity to proteins in the databases are described in 'product' or 'note' qualifiers. Genes that have no significant protein similarity are described as 'unknown protein'.
The software programs used to predict genes include: Grail (Informatics Group, Oak Ridge National Laboratory, http://compbio.ornl.gov/Grail-1.3/), GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html), NetGene2 (S. M. Hebsgaard, et al., CBS, Technical University of Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and SplicePredictor (Volker Brendel, Stanford University, http://gremli.ni.zoel.iastate.edu/cgi-bin/sp.cgi).
Genes encoding tRNAs are predicted by tRNAscan-SE

(Sean Eddy, Washington University School of Medicine, St. Louis, <http://genome.wustl.edu/eddy/trNAscan-SE/>). of this clone. It may be shorter because we remove overlaps between neighboring submissions. The 5' clone is K9122 and the 3' clone is MUL14.

FEATURES

SOURCE

1. .83650
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
/chromosome="3"
/clone="MPE11"
/clone_lib="Mitsui P1"
/ecotype="Columbia"
complement(342..764)
/note="CDS is reported in Acc# AP000599
contains similarity to CHP-rich zinc finger protein
gene id:K9122.5"
/number=1

EXON

CDS
join(1705..2463,2548..2739,2862..2949,3037..3107,
3190..3273,3410..3514)
/notes="gb|AAD55139.1
gene id:MPE11.1"
/codon_start=1
/evidence=not_experimental
/product="dihydroliipoamide S-acetyltransferase"
/protein_id="BAB01047.1"
/db_xref="GI:9279589"
/translation="MTVRGKIRIFMPALSSMTTEGKIVSWIKTEGEKLAKGESVVVV
ESDKADMDVETFDVGLAIVVGEGETAPVGAIGLLAETAEIEBAKSAKSSSS
VBAVVPSPPTVSSAPALAPAPVATVSDGPRKVTATPYAKLAKQKHQVIESVAG
TGFGRITASDVETAGIAPSKSSIAPPPDPPTAKATTNLPPLDPSSIVPPTA
MQSASKNMIESLVFTFRVGVAVNTDALDYKVKPKGVMTALLAKAAGMALAQH
PVNASKDKGFSYNSINIAVAINGLITPVLQDADKLDLYLLSKOKELVGKA
RSKLOPHEVNSTFTLSNLGMFGVDVFDAIIPPGOGAIMVAGSKPTTVADKGFSS
VKNTMLVNTADHRIYVGADLAFLQTFAKLIENPDSLTL"
complement(4594..5106)
/notes="unnamed protein product; gene_id:MPE11.2
unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB01048.1"
/db_xref="GI:9279590"
/translations="MEDLLEERLTRDTSVGNKRVROGLDLDSPDVRLRDDLPDSSGL
DPVSQDLDSVKMSFENELSTTTAALSGETQDPLGYLFEASDDELGLPLPPTQTL
PSCSETVTELVRASDSSEVGLCGFEDHVTVEFGCDLGDGLFFYFDGCLDSGDLF
SWRPEFLPAE"
complement(join(10090..10408,10649..10804,10916..10962))
/note="unnamed protein product; gene_id:MPE11.3
similar to unknown protein
sp|P42744

CDS

/evidence=not_experimental
/protein_id="BAB01049.1"
/db_xref="GI:9279591"
/translations="MMPEKAKYDRQLMYTIQGTLEBASICLNCGPIGSNALKNLVLG
EQNSITGVSGKVLIGDIWKQFHRHAIEQKFSISEGFRDENNTVFORRHQSVFQRL
GONRIAGQTVRPMIDIARRIWASIGRWMSLADRTTYTKEARAITDPFGRLVRLTSL
VDHLDPRRLCDI"
complement(11450..11818)
/notes="unnamed protein product; gene_id:MPE11.4
unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB01050.1"
/db_xref="GI:9279592"
/translations="MNNSLKKEERVEEDNGKSDCNKSKPSTEVVTVTVEEVDPEPKI
LRKHVHATRVAVKGVGAEGLPFSKKRKRSQNLGRNSLDCNGVRDGEFDEINRVGL
QGIGLDLCKPDPDSVLSL"
16981..17979
/notes="gene_id:MPE11.5"
/codon_start=1

CDS

complement(join(33961..34021,34114..34226,34391..34482,
34602..34800))
/note="unnamed protein product; gb|AAF26483.1
gene_id:MPE11.10
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB01056.1"
/db_xref="GI:9279598"
/translation="MKNVMLIIDESNASYDILLIWALENQKDITSESKVIFAKQPQNS

/evidence=not_experimental
/product="AP2 domain transcription factor-like protein"
/protein_id="BAB01051.1"
/db_xref="GI:9279593"
/translation="MAERKKRSSIQTNKPNKKMKKPPQLNHLPLGLSBDLTKMRKL
FVNDPYATDYSSEERSQRKRKYVEIDLPFAAATAQAESESSYCSNNGNSVK
TKISACKVLBSKASPVVGRSSTTVSKPVGRQKWKGAWEIRHPITKVTWLTGY
ETLEQRADAYATKLEFDALAAATSAASVLSNESGSMISASGSDIDDKLVDSFLD
QOAGSEKASDFDADIQIPMGCFIDDSFINACELDFLLTEENNQMLODDYCGID
DLDIIGLCDSGSELPDYDFSDVEIDGLIGTITIDKYAFVDHIAITTPPLAIACP"
join(21893..22063,22330..22386,22578..22783,22894..23227,
23311..23382,23473..23613)
/note="gb|AAF23821.1
gene_id:MPE11.6"
/codon_start=1
/evidence=not_experimental
/product="homocysteine S-methyltransferase AtHMT-1"
/protein_id="BAB01052.1"
/db_xref="GI:9279594"
/translation="WVLEKKSALLEDLIKKCGCAVVDGGFATQLEIHGAANPLWS
AVSLTKNPELIRKRMXYLEAGADIVVTSSYQATIPGFLSRGLSIESSLSLQKSVL
AVEARDRFEKVSQVSGHSYNRALVAAGISYGAYLADGEYSGHYGENVSDKLKDF
HRRRLQVLVEAGPDLAPETIPNKLEAOACVELLEEKVQIPAWICFTSVDEKAPSG
ESFECLEPLKNNIYAVINCAPPQETENLIRKFAKLTKKAIIVVYVNSGEWIDGKA
KOWLPSCQFGDDEPFMPATKWRDLGAKLIGGCCRTTPSTINALSRLKRR"
complement(27277..27552)
/note="unnamed protein product; gene_id:MPE11.7
unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB01053.1"
/db_xref="GI:9279595"
/translation="MTHARERSSLTTLMLVILLSYMLHLFCVYSRVGAIRIFPPTP
ASGKQEDLMKKYFGAGKFPVDSFVGKIGSRIKRVSPSCDPLHN"
37065..31883
/note="unnamed protein product; gb|AAF16548.1
gene_id:MPE11.8
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB01054.1"
/db_xref="GI:9279596"
/translation="MPKRRKPSVSLDKYKRSPLCEASLALKPSEKOVKEWEARCP
VCMERPHGILLICSSYENGCRPYMCDTSHRSNCFDQPRKASKKPSLSLREBES
NPTMEDVDSDTAVNLGEAASETIVVDLSDGBRGEVEEVEEVEVEVEEIGIV
TTEEDQENKPKQLTCPLCRHIKEWVYVAAKCFMNSKXSCCTCDPSGYSDLR
KHARLHPVRPSEADPRQRWRRLEROSLDLSTLQSSFGGDEISNDGFLPAD
TBLLTVPYLRVVRPESGSSSSWSGTSRARTHTSGRRSSRSPASLWGESYEGNTGT
SPRDENNQSSDEQVSGTFRRSRRTVIDDDDEEP"
complement(join(32197..32430,32525..32676,32969..33332,
33404..33487))
/note="gene_id:MPE11.9"
/codon_start=1
/evidence=not_experimental
/product="chloroplast 50S ribosomal protein L15"
/protein_id="BAB01055.1"
/db_xref="GI:9279597"
/translation="MATPLSISNPLTSRHCYRLHLSSTSFKNYSVLGANPFSQILSL
KLNQTLKYNQQQFARPLVVSOTATSSAVAPERLDNLGPOPSRKKOKRKGKRG
ISAGOGASCGFCMGKRSRSGPIMRGEGGOTALYRLPKLGGTAGMWSGLPKYLP
VNIKDIETAGFQEGDEVSLTKQKGLINPNSRERKLPLKILGTGELSMLKTFKARAF
A"
STQAKEUEASGCTLTIVPGRKKVVKPSVAKNQARAEYFAKKRAAAAASEPAAS
complement(join(33961..34021,34114..34226,34391..34482,
34602..34800))
/note="unnamed protein product; gb|AAF26483.1
gene_id:MPE11.10
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB01056.1"
/db_xref="GI:9279598"
/translation="MKNVMLIIDESNASYDILLIWALENQKDITSESKVIFAKQPQNS

CDS
FTPPTVSSVGAQIYPFPSPNSSELIRIAQERNKIALGILEKAKKICLNHGIIKAT
FTDDGPKLIRIKIIOERNINLIVTSDQSLKKCTQNTDCSLIVVKRLKRD
join(36033..36100,36243..36283,36497..36565)..36707,
36797..36871)
/note="unnamed protein product; contains similarity to
transcription-associated zinc ribbon protein
gene_id:MPB11.11"
/codon_start=1
/evidence=not experimental
/protein_id="BAB01057.1"
/db_xref="GI:9279599"
/translation="MEKSRSEBFLPCNLGTMVLVSKTYVAECPHCKTTRNAKDIIDK
EYATVSAEDIRLEGLISUFGKIQAEALPDKKACEKQHPDLVYTRQKRSADG
QTTYTCPCNAHRFTG"
Query Match 100.0%; Score 962; DB 8; Length 83650;
Best Local Similarity 100.0%; Pred No. 9.6e-140; Indels 0; Gaps 0;
Matches 962; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGTGTTCTCAGAAATAGCAGCAAAATATTTATAAAAGCATGCAATTTCTTTATAGATCG 60
DB 12780 TGTGTTCTCAGAAATAGCAGCAAAATATTTATAAAAGCATGCAATTTCTTTATAGATCG 12721
QY 61 CGAAGTTTAAAAAACAATATAGAAATGTTTACAATATACATGGTTTTTATTCGATAACA 120
DB 12720 CGAAGTTTAAAAAACAATATAGAAATGTTTACAATATACATGGTTTTTATTCGATAACA 12661
QY 121 TGACAAATATTTATTTTATTTTCATGAGTTTTTATTTGGATAGCATGACAAATATTAATATAT 180
DB 12660 TGACAAATATTTATTTTATTTTCATGAGTTTTTATTTGGATAGCATGACAAATATTAATATAT 12601
QY 181 CAGTGTAAATACATGTTTGTGTTCTTAAATACATGCAATTTTAAATTCAGACATTTGTTT 240
DB 12600 CAGTGTAAATACATGTTTGTGTTCTTAAATACATGCAATTTTAAATTCAGACATTTGTTT 12541
QY 241 TAAATCAATCTAATCTCTTATATCACAACGACATTCAGGAAATTCAGGTAAAGAA 300
DB 12540 TAAATCAATCTAATCTCTTATATCACAACGACATTCAGGAAATTCAGGTAAAGAA 12481
QY 301 GAAATTAAGAATGAGAGATAGAGATTTCTATGAAAGAAAGAGAGAACATCTAGG 360
DB 12480 GAAATTAAGAATGAGAGATAGAGATTTCTATGAAAGAAAGAGAGAACATCTAGG 12421
QY 361 TGAACAAAATAAAGAGATATGATGATATATTTTATGAGAGGTGGTGAAGATTATTTTAGG 420
DB 12420 TGAACAAAATAAAGAGATATGATGATATATTTTATGAGAGGTGGTGAAGATTATTTTAGG 12361
QY 421 AGAGGAGAGAGAAATAGAAAAGAAATGACATGGTGAATCTGAAAGAGATGAATTTGTG 480
DB 12360 AGAGGAGAGAGAAATAGAAAAGAAATGACATGGTGAATCTGAAAGAGATGAATTTGTG 12301
QY 481 TTAAGATGAAGAGAGAAAGAGAACTCCATGGCTAAAGTCTCGTAAAGAGATGAAAG 540
DB 12300 TTAAGATGAAGAGAGAAAGAGAACTCCATGGCTAAAGTCTCGTAAAGAGATGAAAG 12241
QY 541 AAAACAAAGAGGAGAGAGAGAGAGAGAGCTTAAATAGACTAACTATTGCCAAAATTTTC 600
DB 12240 AAAACAAAGAGGAGAGAGAGAGAGAGAGCTTAAATAGACTAACTATTGCCAAAATTTTC 12181
QY 601 TGTAGCCGCAAAATACATATTTTGTGTCAGAGTTATTTTGTGATTTCTTTTGAAGTCAAAAG 660
DB 12180 TGTAGCCGCAAAATACATATTTTGTGTCAGAGTTATTTTGTGATTTCTTTTGAAGTCAAAAG 12121
QY 661 TTATTTCTTACATATCTCTTAAATATATAGCCGATACCAATTTTCCACACATGGACTTC 720
DB 12120 TTATTTCTTACATATCTCTTAAATATATAGCCGATACCAATTTTCCACACATGGACTTC 12061
QY 721 CTTTATTTCCAAAAGTCAATAAAGTGTGACGTCAATGATCTTACGCTTTTAAACATCCGCAT 780
DB 12060 CTTTATTTCCAAAAGTCAATAAAGTGTGACGTCAATGATCTTACGCTTTTAAACATCCGCAT 12001
QY 781 GATGATGTCATTAGCATCAATCTCCACCGTCCAAATTTATTTAGTTGTTGCAATATCGAC 840
DB 12001 GATGATGTCATTAGCATCAATCTCCACCGTCCAAATTTATTTAGTTGTTGCAATATCGAC 840

DB 12000 GATGATGTCATTAGCATCAATCTCCACCGTCCAAATTTATTTAGTTGTTGCAATATCGAC 11941
QY 841 CGTCTAAGTTCCACACCGCAGCGCTATAAGAGTTTCATTATAAATTTTAGCAAAATAAAAT 900
DB 11940 CGTCTAAGTTCCACACCGCAGCGCTATAAGAGTTTCATTATAAATTTTAGCAAAATAAAAT 11881
QY 901 CAGCAAAATAATTTTTTCTTTGACTAAGCTTTAAACGACCGCTTAAACATTTTCTTCTGGCTA 960
DB 11880 CAGCAAAATAATTTTTTCTTTGACTAAGCTTTAAACGACCGCTTAAACATTTTCTTCTGGCTA 11821
QY 961 AC 962
DB 11820 AC 11819
RESULT 2
AR488147 LOCUS AR488147 1700 bp DNA linear PAT 15-MAY-2004
DEFINITION Sequence 3 from patent US 6706952.
ACCESSION AR488147
VERSION AR488147.1 GI:47253921
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1700)
AUTHORS Cad, R.M. and Dietrich, R.A.
TITLE Arabidopsis gene encoding a protein involved in the regulation of
SAR gene expression in plants
JOURNAL Patent: US 6706952-A 3 16-MAR-2004;
FEATURES
Location/Qualifiers
source 1..1700
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Query Match 89.6%; Score 862; DB 6; Length 1700;
Best Local Similarity 100.0%; Pred. No. 6.8e-124;
Matches 862; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 101 TGGGTTTTTATTTGGATAACATGACAAATATTTATTTATTTTCATGAGTTTTTATTTGGATAG 160
DB 1 TGGGTTTTTATTTGGATAACATGACAAATATTTATTTATTTTCATGAGTTTTTATTTGGATAG 60
QY 161 CATGACAAATATTAAT 220
DB 61 CATGACAAATATTAAT 120
QY 221 TTTAAATCAGACATTTGTTTTTAAATCAAATCTAATCTCTTATATATCAACGACATTGAC 280
DB 121 TTTAAATCAGACATTTGTTTTTAAATCAAATCTAATCTCTTATATCAACGACATTGAC 180
QY 281 GGAAATTCAGGTAAAGAGAGAAATATAAGATAGAGATAGAGATTTCTTATGAAATA 340
DB 181 GGAAATTCAGGTAAAGAGAGAGAAATATAAGATAGAGATAGAGATTTCTTATGAAATA 240
QY 341 AGAAAGAGAGACATCTAGTCTGACAAATAAAGAGATATGATATATATTTTATCAGAG 400
DB 241 AGAAAGAGAGACATCTAGTCTGACAAATAAAGAGATATGATATATTTTATCAGAG 300
QY 401 GTGGTGAAGATTATTTTATGAGAGGAGAGAGAAATAGAAAAAGAAAAATGACATGGTGAA 460
DB 301 GTGGTGAAGATTATTTTATGAGAGGAGAGAGAAATAGAAAAAGAAAAATGACATGGTGAA 360
QY 461 TCTGAAGAGAGATGAATTTGTTTAAAGATGAAGAGAGAGAGAACTCCATGGCTTAAAGTC 520
DB 361 TCTGAAGAGAGATGAATTTGTTTAAAGATGAAGAGAGAGAGAACTCCATGGCTTAAAGTC 420
QY 521 TCGTAAAGAGATGAAAAAGAACAAAG 580
DB 421 TCGTAAAGAGATGAAAAAGAACAAAG 480
QY 581 CTAACCTATTGCCAAAATTTCTGTAGCCGACAAATACTATTTGTGTCACAGGTTATTTTGTG 640

Db	481	CTAACTATTGCGCAAAATTTCTGTAGCGGCAAAATACTATTTTGTGTCCAGGTTATTTTGTG	540
Qy	641	TATCTTTTGAAGTCAAAAGTTATTTCTTACATATACTCTTAAATAATATAGCGGATACCAA	700
Db	541	TATCTTTTGAAGTCAAAAGTTATTTCTTACATATACTCTTAAATAATATAGCGGATACCAA	600
Qy	701	TTTTTCCACATGGAAGTCTCTTTATTTCCAAAGTCAATAAAGTGTGACGTCAATGATACT	760
Db	601	TTTTTCCACATGGAAGTCTCTTTATTTCCAAAGTCAATAAAGTGTGACGTCAATGATACT	660
Qy	761	TAGCGTTTAAACATCGCATGATGTGCTATTAGCATCAATCTCCACCGTCCCAATTTATT	820
Db	661	TAGCGTTTAAACATCGCATGATGTGCTATTAGCATCAATCTCCACCGTCCCAATTTATT	720
Qy	821	TAGTTGTTGACAAATCGACCGTCTAAGTTTCCACACCGCGGTATAAGAGTTTCATTAT	880
Db	721	TAGTTGTTGACAAATCGACCGTCTAAGTTTCCACACCGCGGTATAAGAGTTTCATTAT	780
Qy	881	AAATTTTAGCAAAATAAATCAGCAATAATATTTTCTTGACTTAAGCTTAAACGACGCCG	940
Db	781	AAATTTTAGCAAAATAAATCAGCAATAATATTTTCTTGACTTAAGCTTAAACGACGCCG	840
Qy	941	TTAAACATTTTCTCTCGCTAAC	962
Db	841	TTAAACATTTTCTCTCGCTAAC	862
RESULT 3			
AB026636			
LOCUS	AB026636	92620 bp	DNA linear PLN 14-FEB-2004
DEFINITION	Arabidopsis thaliana genomic DNA, chromosome 3, TAC clone:K14A17.		
ACCESSION	AB026636	BA000014	
VERSION	AB026636.1	GI:4757392	
KEYWORDS			
SOURCE	Arabidopsis thaliana (thale cress)		
ORGANISM	Arabidopsis thaliana		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eutrosids II; Brassicales; Brassicaceae; Arabidopsis.		
REFERENCE	1	Sato, S., Nakamura, Y., Kaneko, T., Katoh, T., Asamizu, E. and Tabata, S. Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence features of the regions of 4,504,864 bp covered by sixty F1 and TAC clones	
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE	2	(bases 1 to 92620)	
AUTHORS	Sato, S., Nakamura, Y., Kaneko, T., Kato, T., Asamizu, E. and Tabata, S.		
TITLE	Direct Submission		
JOURNAL	Submitted (28-APR-1999) Yasukazu Nakamura, Kazusa DNA Research Institute, Department of Plant Gene Research, 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: ynakamu@kazusa.or.jp, Tel: 81-438-52-3935, Fax: 81-438-52-3934)		
COMMENT	Address for correspondence: kao@kazusa.or.jp For the latest information on annotation of this clone, please see http://www.kazusa.or.jp/kaos/cgi-bin/sgd_graph.cgi?c=K14A17 Genes with similarity to proteins in the databases are described in 'product' or 'note' qualifiers. Genes that have no significant protein similarity are described as 'unknown protein' The software programs used to predict genes include: Grail (Informatics Group, Oak Ridge National Laboratory, http://compbio.ornl.gov/Grail-1.3/), GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html), NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and SplicePredictor (Volker Brendel, Stanford University, http://gremli.zool.iastate.edu/cgi-bin/sp.cgi). Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, Washington University School of Medicine, St. Louis, http://genome.wustl.edu/eddy/tRNAscan-SE/). This sequence may not be the entire insert of this clone. It may be		

shorter because we remove overlaps between neighboring submissions.
The 5' clone is MHU15 and the 3' clone is MCE21.

Location/Qualifiers
1. .92620
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
/chromosome="3"
/clone="K14A17"
/ecotype="Columbia"
complement (201..2180)
/note="gene_id:K14A17.1"
/codon_start=1
/evidence=not experimental
/product="salt-inducible protein-like"
/protein_id="BAA94973.1"
/db_xref="GI:7670019"
/translation="MRGFASASRIATAAAASKSLNASTSVNPKLSKTLNSSGKPTNP
LNQYISQVIERKDWFLILNQEFTHRIGLNTFRVISVLQNDQNPULHSURFLWVSNF
DPVYAKDSLSLGNALFRGFLLSMELLKEIRDSGYRISDELMLCVLIGWGRGLG
AKYCNDFQAISFLGKPSRILYNAVIDALVKNSLDLAYLKFQOMRSQCKPDRPM
NLIHGVCCKGVDEAIRLVKOMEQGNRPVFTVILIDGLFIAGRVDEALQKLRPM
RVKLNPEATIRTFVHGIFRCLPCKAFVLVGMKDSNLQVGYDAVLVCLSNNS
MAKETGQFLKIGERGYIPDSSTFNAASCLLKGHDLVETCRIFDGVSRGVKGPGNG
YLVVQALNLAQRSEGRYLKQMGVGLSSVSYNAVIDCLCKARRIENAMFLUTE
MODRGISPLNFTFTLSGYSVRGVGVHLEKLVHGFKPFVITFLLIINCLCRA
KEIKDAFCFEMLEWGEIPNEITYNIRISCCSTGDTDSVKLFAMKENGSLSPDY
AYNATIOSFCMKVKAEELKTLRLIGLKPDNFTYSTLIKALSESREAREMFS
STIERGCVDPDSTKRLVELDLRKSLSPRETVSAS"
join (3906..4124,4238..4462,4536..4649)
/note="unnamed protein product; emb|CAB82745.1
gene_id:K14A17.2
similar to unknown protein"
/codon_start=1
/evidence=not experimental
/protein_id="BAA94974.1"
/db_xref="GI:7670020"
/translation="MAKQDPTSSNTMLPLVGEFVRPQPLDLTITGDTVKDATGNKRVF
KYKTLFGLHNKRILVDPNDSPIVTMKKVTSKHWRQVYRGSDDDKITFTKRSSTV
QLKTRVEVFLKHNQTRSCDFTIKGRFMKRACTIIVADSTKIIAQVYEGHERLVATI
VNVDPYAFITVLIFIPDLINMGDTGI"
complement (join (5031..5927,6011..6601,6984..7211))
/note="gene_id:K14A17.3"
/codon_start=1
/evidence=not experimental
/product="AMP-binding protein"
/protein_id="BAA94975.1"
/db_xref="GI:7670021"
/translation="MAATKWRDIDDLPKIPANYTALTPLWFLDRAAVVPTPKRSVING
SREYTRQTYDCRRLASALADRSIGPGSTVFVAIAPNIPAMYEAHFVPMCGAVLN
CVNIRLNAPTVALLSHSQSSVIMVDQEPFTLAEDSLRLMEKAGSFKRPLLIIVIGD
HTCPESLNRLASKGAI EYEDFLATDPNYPWQPADQMGSTALGYSTGTTASPKGV
LHRRGAI MALSNPLIWMQDCAVYLWTLPMFCHNGCWPMSLAVLSGTSICLRQVTA
KEVYSNIAKYKTHFCAPVLNAINAPKEDTILPHTVHVMTAGAAPPSVLSFM
NQGRPRVATYGLSETYGPSTVCAMKPEWSDUPPETQAKLNARQVRYRMBQDLVID
TQTGKPVADGKTAGEIVFRGNMVKGLKNPEANKETFAGGWSGSDIATVHPDNYI
EIKDRSKDVIISGGENISSEVENYVYHHPVLEASVVARPDQERWSPCAFTVLSKD
YKHKQNLQAQDIMKFCREKLPAYVWPVKSVEFGPLPKTATGKTIQKHILRTKAKEMGPV
PSSRT"
complement (join (8463..8892,8977..9136,9230..9641))
/note="gene_id:K14A17.4"
/codon_start=1
/evidence=not experimental
/product="basic chitinase"
/protein_id="BAA94976.1"
/db_xref="GI:7670022"
/translation="MVSRLPFSLLLTVALVVFQGTGLVNAEDSEPSSTRKPLVKIV
KKGKICDQGWCKGWSYCCNHTIISDFETQYFENLFSKNSPFAHVGWYRSPIT
AAAYQPLGFTAGKLGQMEKVAAPLGHVGSKTSCGYGATGGLGAWLGCYKNKMS
DQLYCDDYYKLTYPCTPGVSGHGRGALPVYNNYNGQTGKALVDLLSHPEYLENNAT
LHFAQAAIWRWMTPPKXHLPSAHDVFGWKPKNTDTAAKRTPFQFATINLVYGDQICN
SGFDNEMNIVSHVLYLIDLVGVRGEEAGPHEKLSQADQEPFSSSSSSSPSSGSSS"

```
CDS
join(13759..13822,14371..14445,14592..14707,14800..14888,
15014..15059,15238..15433,15520..15845,15928..17178,
17360..17713)
/notes="gene_id:K14A17.5"
/codon_start=1
/evidence=not_experimental
/product="transcription factor-like protein"
/protein_id="BAA94977.1"
/db_xref="GI:7670023"
/translations="MDDGLGRIGSIHGFIHLQDLVDVTMLEEAKSRWLNRPEIHA
ILCGRIILPRKMLRPRKDGHNKKKDGRTVKEAHEHLKVGNEIRIHVYVAHGDN
TTRVRCYLLDKARENILVHYRDQEAATTSQDSISSPISVSEQTFFNRVAEEDID
TVNRHDSIHLNDLNDVLLVYRDQEAATTSQDSISSPISVSEQTFFNRVAEEDID
TVADGSLDALLNDPOSRESFGWMMNSFISESGSLSDPSFEPVMVPRQDPLAPQAVF
HSHSNIPQEVNTDVSPPAWAYSEKTKILVTGFLHDSYQHLERSNLYCVCGFCVPA
EYLQAGFYRLLIIPPHSGVMNLYSLADGKPKISQCFRFEHRAVPLDKTVPEDNQSK
WBEFQGVRLSHLLFTSSNKLNLKSLKSPHNLDAKLLASKTNHLLSNWVLYVKSIO
GNKVSFQADKHLFELSLKRLKELWMEKVLGRNTLDVDSKGLVHLCASLGITWS
VQPLSGSLNFRDQGWGTHLWAAVYGRKVMVALLSAGAPNLTVDSTKDNLGGC
MAADLAQONGYDGLAAYLAELKCLVAQFRDMKIAGNITGDLCAKEMLNGTLPDEQ
SLKDAAVYRTAAEAARIQCAFREKALKARSVIOFANKBEEAKSIITAMKIQNAF
RKVDTRKIEAAVRIOCRFPOTWIKIRREYLMNRQAIRIOAARGLQARQYKILQWSV
GVLEKAVLRWKEKGFGLQVAEEDSPGEAQEDFYKTSQRAERLERSVVRVQAM
FRSKKAQDYRMKLTHERAQVNLHFLNLSPGKNSNR"
complement(join(18170..18172,18407..18490,18736..18807,
18977..19090,19168..19281,19387..19487,19590..19668,
19755..19805,19906..19989,20365..20431,20539..20669,
20916..21040,21126..21231,21349..21483,21624..21770,
21860..22171))
/notes="gene_id:K14A17.6"
/pseudo
/codon_start=1
/evidence=not_experimental
/product="lipamide dehydrogenase"
join(29096..29206,29695..29826,29925..30009,30149..30312,
30830..31267)
/notes="unnamed protein product; contains similarity to
similar to ubiquitin conjugating enzyme
gene_id:K14A17.7"
/codon_start=1
/evidence=not_experimental
/protein_id="BAA94978.1"
/db_xref="GI:7670024"
/translations="MADERYRNKNRPAVKRILOVBKQMNQPSDDFMSLPLEENIFWQ
FAIRGPDVFEFGGIVHGIQLPADYPPKPPSMLTNGRPTNTKICLSISNHYPE
HWPQSVSTVALIATFMTSPNGALGSVDYKPKDRRTLAKSRPTPKYSGSPERQK
IIDEIHQYLSKATVPKPLPESQAPSIIVSAHQVPEQBAITVVERSTATTDTI
VDQIITEETAENVTAASVVPAAAPLPAVEVVVKSASVSGEQRMARAAOKPVDRLFT
WAAVGUTIAIMVLLKKFKNSGYSTGFMDQDS"
join(32260..32364,33121..33412,33506..33750,33837..34103)
/notes="unnamed protein product; gb|AAD26962.1
gene_id:K14A17.8
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAA94979.1"
/db_xref="GI:7670025"
/translations="MCKSSNIVEDRIKEBQMLSFKEKIFQADLSDSCMRALPYSFIS
KVEKDFSKYKVIQAWGKTWDEVEKNPTYYIETRGWDLFVSDNALGONFITETH
RGNMVFHVIYEQNGLEMRKPRKFTQMGSSGKKEEGNSLIDVKKKEESDESPORA
ELVLRKKTTEDSKSKKTRNKRKKSQKQVLDGVPEFKITIRKSYLFLAIPK
HFVDDHI PNKSKITIRHPNGSGWKVLCVREIRTFISGYSKLAREFFLWVGDKCTF
KLIPKPEFVLLTSKKNREKWDQCMID"
complement(join(34528..34725,35213..35322,35680..35863))
/notes="unnamed protein product; gb|AAF26101.1
gene_id:K14A17.9
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAA94980.1"
/db_xref="GI:7670026"
/translations="MAESGGRIGGIVAFDSDCKKALSWAIDNVVRDGHLLITIAH
DMNYERGEQJLWETVGSPPFIPMSFSDAAVMKKYALKPDAETLIDIVNTAARKKTIW
```

```
Query Match      70.3%; Score 676.2; DB 8; Length 92620;
Best Local Similarity 99.6%; Pred. No. 1.2e-95;
Matches 678; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGTGTTTCTCAGAAAATAGCAGCAAAATATTTTATAAAGCATGCAATCTCTTATAGATCG 60
DB 71376 TGTGTTTCTCAGAAAATAGCAGCAAAATATTTTATAAAGCATGCAATCTCTTATAGATCG 71435

QY 61 CGAAGTTTAAAAAACAATATAGAAATGTTTACAATATTACATGGGTTTTTATTTCGATAACA 120
DB 71436 CGAAGTTTAAAAAACAATATAGAAATGTTTACAATATTACATGGGTTTTTATTTCGATAACA 71495

QY 121 TGAACAATATTTATTTATTTTCATGAGTTTTTATTTGGATAGCATGACAAAATATTAAATATAT 180
DB 71496 TGAACAATATTTATTTATTTTCATGAGTTTTTATTTGGATAGCATGACAAAATATTAAATATAT 71555

QY 181 CAGTGTTAATAACATGTTTGTTCCTTAAATAACATGCAATTTTAAAAATCAGACATTTGTTT 240
DB 71556 CAGTGTTAATAACATGTTTGTTCCTTAAATAACATGCAATTTTAAAAATCAGACATTTGTTT 71615

QY 241 TAAATCAATCTAATCTCTTATATCACAACGACATTTGACGGAATAATTCAGGTAAAGAAGA 300
DB 71616 TAAATCAATCTAATCTCTTATATCACAACGACATTTGACGGAATAATTCAGGTAAAGAAGA 71675

QY 301 GAAATATAAGATGAGAGATAGAGAGATTTCTATGGAATAAGAGAGAGAAACATGTAGG 360
DB 71676 GAAATATAAGATGAGAGATAGAGAGATTTCTATGGAATAAGAGAGAGAAACATGTAGG 71735

QY 361 TGACAAAATAAGAGATATGATGATATATTTTATGAGAGGTGGTGAAGATTTATTTTAGG 420
DB 71736 TGACAAAATAAGAGATATGATGATATATTTTATGAGAGGTGGTGAAGATTTATTTTAGG 71795

QY 421 AGAGGGAGAGAGAAATAGAAAAGAAATACACATGCTGTAATCTGGAAGAGATGAATTTGTG 480
DB 71796 AGAGGGAGAGAGAAATAGAAAAGAAATACACATGCTGTAATCTGGAAGAGATGAATTTGTG 71855

QY 481 TTAAGATGAAGAGAGAGAAAGAGAACTCCATGGCTAAAGTCTCGTAAAGAGATGAAAAAG 540
DB 71856 TTAAGATGAAGAGAGAGAAAGAGAACTCCATGGCTAAAGTCTCGTAAAGAGATGAAAAAG 71915

QY 541 AAACAAAAGAGAGAGAAAGAGAAAGCTTAAATAGACTAACTATTTGCCCCAAAATTTTC 600
DB 71916 AAACAAAAGAGAGAGAAAGAGAAAGCTTAAATAGACTAACTAACTATTTGCCCCAAAATTTTC 71975

QY 601 TGTAGCCGCAATAACTATTTCGTCCCAAGTTATTTTGTCTATTCTTTTGAAGTCAAAAG 660
DB 71976 TGTAGCCGCAATAACTATTTCGTCCCAAGTTATTTTGTCTATTCTTTTGAAGTCAAAAG 72035

QY 661 TTATTTCTTTACATATACTCTA 681
DB 72036 TTATTTCTTTACATATCTCA 72056

RESULT 4
AB005248/c
LOCUS
Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone:MX110.
DEFINITION
AB005248 BA000015
ACCESSION
AB005248.1 GI:2264320
VERSION
KEYWORDS
SOURCE
Arabidopsis thaliana (thale cress)
```

ORGANISM	Arabidopsis thaliana	/note="unnamed protein product; gene_id:MXI10.2
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi	pir T03994
AUTHORS	1 Sato,S., Kotani,H., Nakamura,Y., Kaneko,T., Asamizu,E., Fukami,M., Miyajima,N. and Tabata,S.	/codon_start=1
TITLE	Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence features of the 1.6 Mb regions covered by twenty physically assigned P1 clones	/evidence=not experimental
JOURNAL	DNA Res. 4 (3), 215-230 (1997)	/protein_id="BAB09347.1"
MEDLINE	97471969	/db_xref="GI:9758813"
PUBMED	9330910	/translation="MWASSLSPSPCLSYFKMPIAALNSPRTTTPYNLKKQTRLHLKPLSGKMEVIEQRKSKSERENPPLVFGHSGYHAAWENMLPFSSGFGDSYAVSLQSGSDPLGTAGTQTQTHASDIADFIESNLGSSPPVLVGHSGFGLIVQYLANIVNKRSLGTENAFPELSGAVMVCSVPPSGNSGLVRLYFSKPVAAFKVLTLSLAAGFQKSIPLCRETFSQAMDDQLVKRYNLAR"
REFERENCE	2 (bases 1 to 83646)	join(7788..7835,8043..8777,9032..9157,9205..9901,10085..12958)
AUTHORS	Nakamura,Y.	/note="gene_id:MXI10.3"
TITLE	Direct Submission	/pseudo
JOURNAL	Submitted (02-JUL-1997) Yasukazu Nakamura, Kazusa DNA Research Institute, Department of Plant Gene Research, 1532-3, Yana, Kisarazu, Chiba 252-0812, Japan (E-mail:ynakamu@kazusa.or.jp, Tel:81-438-52-3935, Fax:81-438-52-3934)	/codon_start=1
COMMENT	Address for correspondence: kaos@kazusa.or.jp For the latest information on annotation of this clone, please see http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=MXI10 Genes with similarity to proteins in the databases are described in 'product' or 'note' qualifiers. Genes that have no significant protein similarity are described as 'unknown protein'. The software programs used to predict genes include: Grail (Informatics Group, Oak Ridge National Laboratory, http://combio.ornl.gov/Grail-1.3/), GENSCAN (Chris Burge MIT, http://CCR-081.mit.edu/GENSCAN.html), NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and SplicePredictor (Volker Brendel, Stanford University, http://grenlin.zool.iastate.edu/cgi-bin/sp.cgi). Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, Washington University School of Medicine, St. Louis, http://genome.wustl.edu/eddy/tRNAscan-SE/). This sequence may not be the entire insert of this clone. It may be shorter because we remove overlaps between neighboring submissions. The 5' clone is MG17 and the 3' clone is MB18. Location/Qualifiers 1..83646 /organism="Arabidopsis thaliana" /mol_type="genomic DNA" /db_xref="taxon:3702" /chromosome="5" /clone="MXI10" /clone_lib="Mitsui_P1" /ecotype="Columbia" join(1791..2912,3016..3318,3584..4660) /notes="gene_id:MXI10.1" /codon_start=1 /evidence=not experimental /product="disease resistance protein-like" /protein_id="BAB09346.1" /db_xref="GI:9758812"	/evidence=not experimental /product="retroelement pol polyprotein-like" 32202..33620 /note="unnamed protein product; contains similarity to retroelement pol polyprotein gene_id:MXI10.7" /codon_start=1 /evidence=not experimental /protein_id="BAB09350.1" /db_xref="GI:9758816" /translation="WVEGTRTNPKRKTWSTEQSPVVAHTWKEDKALKVLSPLSPFTPAFKREILKRGKLTIDEFTKMGGLAFTSMDFTEPTVDFSTWEYTFITNPKPIAKAGIINFVKTKAFSIPDLCEAYGFENKSMSPFKGIHFFWGLASGK FSCNASKIRHVPVIRYALKLHALFGRGETSSTSEMCFLOQVKELLVEDADEN GNRPLVHDLDGNVYGCIFASYLDKYNRLIKGKTIYIGCTITPLPVKAGVDLSP FNALPKREIDYENLVRSCLKRSDDAFILVTDREDNKLECILPSKETTNVESHED I MFLPSKEQVFNLEQFMEIEEDLPPSRDDDPYNLKNKCDLPSTFPVPTKMEKFLQSV IKSQIKNKNSRYAIGKLLKVKVKLPQPDYVSEDELFPSPNEEDRDNDILSDENNE TPATSSHSRRKNASRDHSYMP"
FEATURES	source	join(18792..18841,19220..19412,19494..19579,19875..19959,20464..20607,20697..20783,20894..21006,21097..21162,21216..21308,21350..21509) /note="unnamed protein product; gene_id:MXI10.5 unknown protein" /codon_start=1 /evidence=not experimental /protein_id="BAB09349.1" /db_xref="GI:9758815" /translation="NMSPKQISDDRGSSHFHRTFPQIIHIGNFRIMSVSMYMYRLN OTGAPVLPFCCLVPSIIIFLIIOKPMKGRALSNOQIVPSLINGVITAYLILWKGK LKSCGPLRALILSEYSGVLGVLGVYGRGHVWKLPKTDSTKEEVRQEQALGM MQMMPVAGILSALRRVIARRVSLKNQKKELHAITTSATCFLPVAMWDLIIGSS SGTSELSPSAWAFLLTIIFGILLIFYDNTAERLHMVFPSSFRHLMVAGVCLIVMDI RVNIRGNISRTQEQLFVEIRRSIKRNPSCSIFPYTNDVLLNLVLDVLDVNLSGFIR FSIYPSPFSCFLGEAYANILV" join(26073..30268,30416..31552) /note="gene_id:MXI10.6" /pseudo /codon_start=1 /evidence=not experimental /protein_id="BAB09349.1" /db_xref="GI:9758815" /translation="NMSPKQISDDRGSSHFHRTFPQIIHIGNFRIMSVSMYMYRLN OTGAPVLPFCCLVPSIIIFLIIOKPMKGRALSNOQIVPSLINGVITAYLILWKGK LKSCGPLRALILSEYSGVLGVLGVYGRGHVWKLPKTDSTKEEVRQEQALGM MQMMPVAGILSALRRVIARRVSLKNQKKELHAITTSATCFLPVAMWDLIIGSS SGTSELSPSAWAFLLTIIFGILLIFYDNTAERLHMVFPSSFRHLMVAGVCLIVMDI RVNIRGNISRTQEQLFVEIRRSIKRNPSCSIFPYTNDVLLNLVLDVLDVNLSGFIR FSIYPSPFSCFLGEAYANILV" join(26073..30268,30416..31552) /note="gene_id:MXI10.6" /pseudo /codon_start=1 /evidence=not experimental /product="retroelement pol polyprotein-like" 32202..33620 /note="unnamed protein product; contains similarity to retroelement pol polyprotein gene_id:MXI10.7" /codon_start=1 /evidence=not experimental /protein_id="BAB09350.1" /db_xref="GI:9758816" /translation="WVEGTRTNPKRKTWSTEQSPVVAHTWKEDKALKVLSPLSPFTPAFKREILKRGKLTIDEFTKMGGLAFTSMDFTEPTVDFSTWEYTFITNPKPIAKAGIINFVKTKAFSIPDLCEAYGFENKSMSPFKGIHFFWGLASGK FSCNASKIRHVPVIRYALKLHALFGRGETSSTSEMCFLOQVKELLVEDADEN GNRPLVHDLDGNVYGCIFASYLDKYNRLIKGKTIYIGCTITPLPVKAGVDLSP FNALPKREIDYENLVRSCLKRSDDAFILVTDREDNKLECILPSKETTNVESHED I MFLPSKEQVFNLEQFMEIEEDLPPSRDDDPYNLKNKCDLPSTFPVPTKMEKFLQSV IKSQIKNKNSRYAIGKLLKVKVKLPQPDYVSEDELFPSPNEEDRDNDILSDENNE TPATSSHSRRKNASRDHSYMP"
CDS		join(36310..37248,37409..37549,37633..37824,38363..38519) /note="emb CAB86928.1 gene_id:MXI10.8 similar to unknown protein"

```
/pseudo
/evidence=not experimental
join(39564, 40460, 40627, 40767, 40853, .41185)
/notes="unnamed protein product; gb|AAD55600.1
gene_id:MX110.9
strong similarity to unknown protein"
/codon_start=1
/evidence=not experimental
/protein_id="BAB09351.1"
/db_xref="GI:9758817"
/translators="MOLSYSPELLECHILSFLTTKEALTSVLSKRRNLIAFVNV
DIVDHIRELFMFVRVLAQGNPINKFSLDCSGVDSRDVNCIIONVMVRGVSBN
LSFVDSFEDDNLFPKVFENKVLKGLSYISWLDGSLFLEMLKTLVLESVLGVE
KPFILHAIIPALSELVNNIYWKERDVVDLIVSESLKTLTINFI VCTHTLSFDPIL
AVLSYGVVVDVVEAKMNLPEARISLVVEGDISRVRLINNDLLEDDEYDVLQPE
NVWKLNGIRNIRCLVSLNLEVLISLCCESMPVFNKLSLSIKSRENKQWAMPVLL
RNCPHLETVLVLEGLLHHVTDKGDACDVSREDKRSLSCPVKVLEIKGFQGTIKEM
NMIKHLEYPFPCLEKELKMYEENDPAQLRVPEVITGEMMERYKKSSSCNVQLLVGKL
YKWTTP"
join(42817, 43766, 43839, .43979, 44083, .44403)
/notes="emb|CAB86928.1
gene_id:MX110.10
similar to unknown protein"
/pseudo
/codon_start=1
/evidence=not experimental
join(45089, 45677, 46843, .47158, 47312, .47452, 47544, .47871)
/notes="contains similarity to non-LTR retroelement reverse
transcriptase
gene_id:MX110.11"
/pseudo
/codon_start=1
/evidence=not experimental
/protein_id="BAB09352.1"
/db_xref="GI:9758818"
/translators="MIVDSMKRVARVAVEGKERNSSNGEETRRSQSHARTAKAR
MNYVKSMPPLPKPKSLMIISLWKEDQDQNNVTVEKEKEKEKGTQVNTIRE
NAMLSKCVCEWPHMLTYEEETWFAFGWQVRYDAVMSGKSWFN"
complement(join(50633, .50872, 51043, .51177, 51268, .51438))
/notes="gene_id:MX110.13
splp10798"
/codon_start=1
Query Match 65.7%; Score 632.2; DB 8; Length 83646;
Best Local Similarity 93.8%; Pred. No. 7.8e-89;
Matches 669; Conservative 0; Mismatches 43; Indels 1; Gaps 1;
QY 1 TGTGTTTCTCAGAAATAGCAGAAATATTATATAAAGCATGCAATCTCTTATAGATCG 60
Db 23871 TGTGTTTCTCAGAAATAGCAGAAATATTATATAAAGCATGCAATCTCTTATAGATCG 23812
QY 61 CGAAGTTTAAARAACATATAGATCTTACATATATACATGGTGTATTCGATAACA 120
Db 23811 CGAAGTTTAAARAACATATAGATCTTACATATATACATGGTGTATTCGATAACA 23752
QY 121 TGACAAATATTTATTTTATTTTCATGAGTTTATTTTGGATAGCATGACAAATATTAAATAT 180
Db 23751 TGACAAATATTTATTTTATTTTCATGAGTTTATTTTGGGATAGCATGACAAATATTAAATAT 23692
QY 181 CAGTGTATAACATGTTTGTGTTCTTAAATAACATGCAATTTTAAATCAGACATTTGTTT 240
Db 23691 CAGTGTATAACATGTTTGTGTTCTTAAATAACATGCAATTTTAAATCAGACATTTGTTT 23632
QY 241 TAAATCAATCTAAATCTCTTATATCACACGACATTCAGCGAAATTCAGGTAAAGA 300
Db 23631 TAAATCAATCTAAATCTCTTATATCACACGACATTCAGCGAAATTCAGGTAAAGA-A 23573
```

```
QY 301 GAAATAAAGAAATGAGAGATGAGAGATTTCTATGAAAGAAAGAGAGAAATGTAGG 360
|||||
Db 23572 TAAATAAAGAAATGAGAGATGAGAGATTTCTATGAAAGAAAGAGAGAAATGTGGG 23513
QY 361 TGAACAAAATAAAGAGATATGATGATATATTTTATGAGAGGTGGTGAAGATTATTTTAGG 420
|||||
Db 23512 TGAACAAAATAAAGAGATATGATGATATATTTTATGAAAGGTGGTGAAGATTATTTTAGG 23453
QY 421 AGAGGAGAGAGAAATAGAAAAAGAAATGACATGTTGAATCTGAAAGAGATGAATTTGTG 480
|||||
Db 23452 AGAGGAGAGAGAAATAGAAAAAGAAATGACATGTTGAATCTGAAAGAGATGAATTTGTG 23393
QY 481 TTAAGATGAGAGAGAGAAAGAGAACTCCATGGCTGAAGTCTCGTAAAGAGATGAAAAAG 540
|||||
Db 23392 TTAAGATGAGAGAGAGAAAGAGAACTCTATGACTAAAGTCTCGTAAAGAGATGAAAAAG 23333
QY 541 AAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
|||||
Db 23332 AAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 23273
QY 601 TGTAGCCGACAAATACTATTGTTGTCGAAGTTATTTTGTGTTATTTCTTTTGAAGTCAAAAG 660
|||||
Db 23272 TGTAGCCGACAAATACTATTGTTGTCGAAGTTATTTTGTGTTATTTCTTTTGAAGTCAAAAG 23213
QY 661 TTATTTCTTACATATCTCTAAATAATATAGCCGATACCAATTTTCCACACAT 713
|||||
Db 23212 TTATTTCTTACATATCTCTACATGTTTGTGTTTAAATTTACTTATTTTAAATAT 23160
RESULT 5
AC012394 94487 bp DNA linear PLN 12-OCT-2000
LOCUS Arabidopsis thaliana chromosome 1 BAC F15M4 genomic sequence,
DEFINITION complete sequence.
ACCESSION AC012394
VERSION AC012394.3 GI:6554469
KEYWORDS HTG.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsia.
REFERENCE 1 (bases 1 to 94487)
AUTHORS Lin,X., Kaul,S., Town,C.D., Benito,M.-I., Creasy,T.H., Haas,B.J.,
Wu,D., Maiti,R., Ronning,C.M., Koo,H., Fujii,C.Y., Utterback,T.R.,
Barnstead,M.E., Bowman,C.L., White,O., Niernan,W.C. and Fraser,C.M.
Arabidopsis thaliana chromosome 1 BAC F15M4 genomic sequence
Unpublished
REFERENCE 2 (bases 1 to 94487)
AUTHORS Lin,X. and Kaul,S.
TITLE Direct Submission
JOURNAL Submitted (27-OCT-1999) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org
REFERENCE 3 (bases 1 to 94487)
AUTHORS Lin,X.
TITLE Direct Submission
JOURNAL Submitted (10-DEC-1999) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
REFERENCE 4 (bases 1 to 94487)
AUTHORS Town,C.D. and Kaul,S.
TITLE Direct Submission
JOURNAL Submitted (12-SEP-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
REFERENCE 5 (bases 1 to 94487)
AUTHORS Town,C.D. and Kaul,S.
TITLE Direct Submission
JOURNAL Submitted (05-OCT-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
REFERENCE 6 (bases 1 to 94487)
AUTHORS Town,C.D. and Kaul,S.
TITLE Direct Submission
JOURNAL Submitted (12-OCT-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
```

COMMENT

On Dec 10, 1999 this sequence version replaced gi:6143858.
Address all correspondence to: at@tigr.org

BAC clone F15M4 is from Arabidopsis thaliana chromosome 1
The orientation of the sequence is from SP6 to T7 end of the BAC
clone.
Genes were identified by a combination of several methods: Gene
prediction programs including Genscan+ (Chris Burge,
<http://CCR-081.mit.edu/GENSCAN.html>), GeneMarkHM (Mark Borodovsky,
<http://genemark.biology.gatech.edu/GeneMark/>), GlimmerA (a variant
of GlimmerM, see Mihaela Pertea,
<http://www.tigr.org/softlab/glimmerm.htm/glimmerm.html>, and
GeneSplicer (Mihaela Pertea and Steven Salzberg, contact
mpertea@tigr.org), searches of the complete sequence against a
peptide database and the plant EST database at TIGR
(<http://www.tigr.org/db/tgi.shtm>). Annotated genes are named to
indicate the level of evidence for their annotation. Genes with
similarity to other proteins are named after the database hits.
Genes without significant peptide similarity but with EST
similarity are named as unknown proteins. Genes without protein
or EST similarity, that are predicted by more than two gene
prediction programs over most of their length are annotated by
hypothetical proteins. Genes encoding tRNAs are predicted by
tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>).
Simple repeats are identified by RepeatMasker (Arian Smit,
<http://ftp.genome.washington.edu/RN/RepeatMasker.html>).

FEATURES

source

1. 94487

/organism="Arabidopsis thaliana"

/mol_type="genomic DNA"

/cultivar="Columbia"

/db_xref="taxon:3702"

/chromosome="1"

/clone="F15M4"

/complement(967..1017)

/rpt_family="AT rich"

/complement(1289..1338)

/rpt_family="AT rich"

/complement(4053..4080)

/rpt_family="AT rich"

/complement(6943..9510)

/gene="F15M4.1"

/note="identical to hydroxy methylglutaryl CoA reductase

(AA 1-592) [GB:X15032 [Arabidopsis thaliana]]

complement(701<9493..7515,7607..7953,8115..8296,

8522..>9510)

/gene="F15M4.1"

complement(1019(7255..7515,7607..7953,8115..8296,

8522..9510)

/gene="F15M4.1"

/codon_start=1

/product="hydroxy methylglutaryl CoA reductase (AA 1-592) ;

9510-7255"

/protein_id="AAF16652.1"

/db_xref="GI:6554470"

/translation="MDLRKPKPPVNTNNNGSPRSYQPTSDDDHRRRTATTIAPP
PKASDALPLPLNTAVFFLTFSSVAYLLHWRDKIRYNTPLHVVITITELGAIILI
ASFIYLLGFGDFVQSFISSRASGAWDLADTIDDDHRLVTCSPPTPIVSGAKLPNP
EPIVLEDEEBEIVKSVIDGIVPSLESRLGDCCKRAISRRELQRTVSKSIEGL
PLDGFYBSILGQCCEMPVGIQIPVGIAGPILLDGYEYSVPMATTEGCLVASTNRGC
KAMFISGGATSVLKDGMRAPVPVFPASARRASELKFPLENPFDTLLAVFNRSRP
ARLQSVKCTIAGNATYRFFCCSTGDAGNMVSKGVNVLEYLTDFFDPMDDVIGISGN
FCSDDKPAANVTIEGRKSVCEAVIRGBIVNKVLKTSVAALVELNMLNLTAGSVAE
SLGGFNASNTVIAVFATGQPAQNVESSCITMEAINDKDIIHISVTWPSIEVG
TVGGGTQLASQACNLNKGKATSPESGPNARRLATIVAGAVLAGELSLMSAIAAGQ
LVRSIMKNRSDISGATTTTITTTT"

complement(7258..7283)

/rpt_family="(CAA)n"

complement(11154..11209)

/rpt_family="AT rich"

complement(11551..11580)

/rpt_family="AT rich"

complement(11933..12004)

gene

/rpt_family="AT rich"
12390..13268

/gene="F15M4.2"

/note="similar to unknown protein GB:AAD30603 [Arabidopsis
thaliana]"

join(<12390..12391,12491..12629,12730..12861,12929..13019,
13099..>13268)

/gene="F15M4.2"

join(12390..12391,12491..12629,12730..12861,12929..13019,
13099..13268)

/gene="F15M4.2"

/codon_start=1

/product="unknown protein; 12390-13268"

/protein_id="AAF16653.1"

/db_xref="GI:6554471"

/translation="MFDVKALYVGKELWRETLQSGSRVYKQLGKSNWYEVKISY
PASIPARVYLOLLKNHEMLKINOMRLLTEKLIFKAESLKEVKNKAGLNLVLTLEP
EGIVAIPINSRRYFIYINIVCEEQLMGIPYSWSVLLVALCLVVSFIVPRSLPSLL
TKQQLRSRSHRGKDS"

14056..15506

/gene="F15M4.3"

/note="similar to cinnamoyl-CoA reductase GB:CAA56103

[Eucalyptus gunnii]"

join(<14056..14170,14277..14431,14516..14701,14819..14990,
15096..15285,15362..>15506)

/gene="F15M4.3"

join(14056..14170,14277..14431,14516..14701,14819..14990,
15096..15285,15362..15506)

/gene="F15M4.3"

/codon_start=1

/product="putative cinnamoyl-CoA reductase; 14056-15506"

/protein_id="AAF16654.1"

/db_xref="GI:6554472"

/translation="NAVKQKVCVTGAGGFIASWLKFLLSRGYTVHGTVRDPCDEKND
HLRLDNASKNLKFKADLFDDDEGLFSAIDGSGVFHIASVPFEGEELIKPALTYTK
NVLEACTETKQKVQVVSIAAVYNPKQPDVAKEDCWSDTQVLSLEGWRYYL
AKTLTERALESKKNFADVTLCPSVIIGPRQLQSTLSSSSGLLKFIKGGIKSLUSD
ELYLDVRDADALLLVNREATGYICNSHSLYTDLSLMEKLNKMKYPRKNPESFTE
VKEKEVRPLSAEKLKNGKFRPLEETIDDSVVSFAAGDLPKA"

16281..16361

/rpt_family="(CAT)n"

16955..18423

/gene="F15M4.4"

/note="contains Pfam profile: PF00076 RNA recognition

motif. [a.k.a. RRM, RBD, or RNP domain]"

join(<16955..17161,17385..17529,17639..17755,17844..17884,
17978..18215,18332..>18423)

/gene="F15M4.4"

join(16955..17161,17385..17529,17639..17755,17844..17884,
17978..18215,18332..18423)

/gene="F15M4.4"

/codon_start=1

/product="putative RNA-binding protein; 16955-18423"

/protein_id="AAF16655.1"

/db_xref="GI:6554473"

/translation="MAYQPVGSGPHYLNPSPGDTTFTKVFVGLAWETQSETLRQHF
EQYGIILAVVIADNTGRSGYGFVTFDPAARRACADPTPIIDGRANCNLASLG
RPRPLPVAIVPNMPLDLPRTLMCRVLEHLEIRISNHFHITTSKELSLIMVTAY
GPYVYSSQGLSPYMGQOVLQVGVFAGVNSPVYQGLSQTIPNGHGYTAVQVGS
VPGSHILQLGFTVSTMTSSMPALQAPYPSGIPGPAPVQSHIIVHSQFMQSTASDQ
TTR"

complement(19352..20843)

/gene="F15M4.5"

complement(19352..19378,19455..19499,19581..19699,
19777..19852,19942..20024,20291..20356,20433..20536,
20615..20690,20777..>20843))

/gene="F15M4.5"

complement(19352..19378,19455..19499,19581..19699,
19777..19852,19942..20024,20291..20356,20433..20536,
20615..20690,20777..20843))

/gene="F15M4.5"

/codon_start=1

/product="unknown protein; 20843-19352"

gene

complement(19352..20843)

/gene="F15M4.5"

complement(19352..19378,19455..19499,19581..19699,
19777..19852,19942..20024,20291..20356,20433..20536,
20615..20690,20777..>20843))

/gene="F15M4.5"

complement(19352..19378,19455..19499,19581..19699,
19777..19852,19942..20024,20291..20356,20433..20536,
20615..20690,20777..20843))

/gene="F15M4.5"

/codon_start=1

/product="unknown protein; 20843-19352"

mRNA

complement(19352..19378,19455..19499,19581..19699,
19777..19852,19942..20024,20291..20356,20433..20536,
20615..20690,20777..>20843))

CDS

complement(19352..19378,19455..19499,19581..19699,
19777..19852,19942..20024,20291..20356,20433..20536,
20615..20690,20777..20843))

/gene="F15M4.5"

/codon_start=1

/product="unknown protein; 20843-19352"

repeat_region

complement(7258..7283)

repeat_region

complement(11154..11209)

repeat_region

complement(11551..11580)

repeat_region

complement(11933..12004)


```
/protein_id="AAF1656.1"
/db_xref="GI:6554474"
/translation="MAAISPWLSQSFNSPRVTITDSRCSISAAISVLDSNBEQ
HRTSSRDHVKMRDVMLOJASSVFLPLAISPAFAETNASEAFRVVYDTNKFELSI
POYSIAITGIGPDFTNRESFGKVEAFETLVSLGRDQKVPVGTAKLIDSRASKGFI
YIETLONPEARKHLYSALGMATNGWLNKLYITVGQFTDESAEAOSSKIQKTVKSPF
FI"
repeat_region complement(20998..21026)
repeat_region rpt_family="AT_rich"
gene complement(21140..21167)
rpt_family="AT_rich"
22462..22509

Query Match 61.7%; Score 593.8; DB 8; Length 94487;
Best Local Similarity 94.7%; Pred. No. 6.4e-83;
Matches 641; Conservative 0; Mismatches 22; Indels 14; Gaps 2;

QY 1 TGTGTTTCTCAGAAATAGCAGAAATATTTATATAAAGCATGCAATTCCTCTTATAGATCG 60
Db |||||
QY 1830 TGTGTTTCTCAGAAATAGCAGAAATATTTATATAAAGCATGCAATTCCTCTTATAGATCG 1889
Db |||||
QY 61 CGAAGTTTAAAAAACAATATAGATTGTTACAATATTACATGGGTTTTTTATTCGATAACA 120
Db |||||
QY 1890 CGAATTTAAAAA-----TTACAATATTACATGGGTTTTTTATTCGATAACA 1936
Db |||||
QY 121 TGACAAATATTTATTTTTCATGAGTTTTTATTCGATAGCATGACAAATATTAAATATAT 180
Db |||||
QY 1937 TGACAAATATTTATTTTTCATGAGTTTTTGTGGATAGCATGACAAATATTAAATATAT 1996
Db |||||
QY 181 CAGTGTAAATACATGTTTTGTTCTTAAATACATGCAATTTTAAATTCAGACATTTGTTT 240
Db |||||
QY 1997 TAGTGTAAATACATGTTTTGTTCTTAAATACATGCAATTTTAAATTCAGATTTTGT 2056
Db |||||
QY 241 TAAATCAAAATCAATCTCTTATATCAACAGCATTCAGCGAAATTCAGGTAAAGA 300
Db |||||
QY 2057 TAAATCAAAATTTAACTCTTATATCAACAGCATTCAGCGAAATTCAGGTAAAGA 2116
Db |||||
QY 301 GAAATTAAGAATCAGAGATAGAGATTTCTATGGAAGAAAGAGAGAAACATGTAGG 360
Db |||||
QY 2117 GAAATTAAGAATCAGAGATAGAGATTTCTATGGAAGAAAGAGAGAAACATGTGG 2176
Db |||||
QY 361 TGAAC-AAAATAAGAGATATGATATATTTTATGAGAGGTGGTGAAGATATTTTAG 419
Db |||||
QY 2177 TGAACAAAATAAAGAGATATGATATATTTTATGAGAGGTGGTGAAGATATTTTAG 2236
Db |||||
QY 420 GAGAGGAGAGAGAAATAGAAAAGAAATGACATGCTGATCTGAAGAGATGAATGT 479
Db |||||
QY 2237 GAGAGGAGAGAGAAATAGAAAAGAAATGACATGCTGATCTGAAGAGATGAATGT 2296
Db |||||
QY 480 GTTAAAGATCAAGAGAGAAAGAGAACTCCATGGCTTAAAGTCTCGTAAAGAGATGAAAA 539
Db |||||
QY 2297 GTTAAAGATCAAGAGAGAAAGAGAACTCTATGCTTAAAGTCTCGTAAAGAGATGAAAA 2356
Db |||||
QY 540 GAAACAAAAGAGAGAGAGAAAGAGAAAGGCTAAATATAGACTAATTCGCCAAATTT 599
Db |||||
QY 2357 GAAATTAAGAAGAGAGAGAAAGAGAAAGGCTAAATATAGACTAATTCGCCAAATTT 2416
Db |||||
QY 600 CTGTAGCCGACAATACTATTTCGTCGAAGTTATTTTGTGTTATCTTTTGAAGTCAAAA 659
Db |||||
QY 2417 CTGTAGCCGACAAGTACTATTTCGTTCTAAGTTATTTTGTGTTATCTTTTCAACTCAAG 2476
Db |||||
QY 660 GTTATTTCTTACATATA 676
Db |||||
QY 2477 GTTATTTCTTACATATA 2493
Db |||||
```

```
RESULT 6
AC015450/c AC015450 100806 bp DNA linear PLN 15-JUN-2001
LOCUS Arabidopsis thaliana chromosome 1 BAC F14G6 genomic sequence,
DEFINITION complete sequence.
AC015450
AC015450
AC015450.5 GI:12323968
```

HTG.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopses;
1 (bases 1 to 100806)
AUTHORS Lin,X., Kaul,S., Town,C.D., Benito,M.-I., Creasy,T.H., Haas,B.J.,
Wu,D., Maiti,R., Ronning,C.M., Koo,H., Fujii,C.Y., Utterback,T.R.,
Barnstead,M.E., Bowman,C.L., White,O., Nierman,W.C. and Fraser,C.M.
Arabidopsis thaliana chromosome 1 BAC F14G6 genomic sequence
Unpublished
2 (bases 1 to 100806)
TITLE Lin,X. and Kaul,S.
JOURNAL Direct Submission
AUTHORS Submitted (16-NOV-1999) The Institute for Genomic Research, 9712
TOWN, C.D. and Kaul, S.
JOURNAL Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org
3 (bases 1 to 100806)
TITLE Town,C.D. and Kaul, S.
JOURNAL Direct Submission
AUTHORS Submitted (19-JAN-2001) The Institute for Genomic Research, 9712
TOWN, C.D. and Kaul, S.
JOURNAL Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
COMMENT On Jan 19, 2001 this sequence version replaced gi:12280761.
Address all correspondence to:at@tigr.org

BAC clone F14G6 is from Arabidopsis thaliana chromosome 1
The orientation of the sequence is from SP6 to T7 end of the BAC
clone.
Genes were identified by a combination of several methods: Gene
prediction programs including Genscan+ (Chris Burge,
http://CCR-081.mit.edu/GENSCAN.html), GeneMarkHM (Mark Borodovsky,
http://genemark.biology.gatech.edu/GeneMark/), GlimmerA (a variant
of GlimmerM, see Mihaela Pertea,
http://www.tigr.org/softlab/glimmerm.htm/glimmerm.html, and
GeneSplicer (Mihaela Pertea and Steven Salzberg, contact
mpertea@tigr.org), searches of the complete sequence against a
peptide database and the plant EST database at TIGR
(http://www.tigr.org/tdb/tgi.shtml). Annotated genes are named to
indicate the level of evidence for their annotation. Genes with
similarity to other proteins are named after the database hits.
Genes without significant peptide similarity but with EST
similarity are named as unknown proteins. Genes without protein
or EST similarity, that are predicted by more than two gene
prediction programs over most of their length are annotated as
hypothetical proteins. Genes encoding tRNAs are predicted by
tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
Simple repeats are identified by RepeatMasker (Arian Smit,
http://ftp.genome.washington.edu/RM/RepeatMasker.html).

FEATURES
Location/Qualifiers
1..100806
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="1"
/clone="F14G6"
misc_feature 1
/note="35000 nt before this point were not included in
the submitted sequence, due to overlap with another BAC
[F15M4]"
5217..5774
/gene="F14G6.1"
/note="contains Pfam profile: PF00097 Zinc finger, C3HC4
type (RING finger)"
5217..5774
/gene="F14G6.1"
5217..5774
/gene="F14G6.1"
/codon_start=1
/product="putative RING zinc finger protein; 5217-5774"
/protein_id="AAG51946.1"
/db_xref="GI:12323975"
/translation="MARLLFRLLQEANSTSPAEAPFPNSDLVLILAVLLCALTCIIG

```
gene
LIAVSRCAMLRRIASRRSDTHPPPPVAANKGLKKKVLRLSPKLTYSPPDSPPAEKLV
ECAICLTERAAGDELRLVLCQCGHFVSCIDTWLGSSSHSCPCSRQILVVTRCHKCGGL
PGSSSGPBPDRIRKQEDGPNLP"
/complement (8331. .10137)
/genes="F14G6.2"
/notes="N-term similar to N-term of NAM GB:CAA63101
[Petunia x hybrida] (apical meristem formation), CUC2
GB:BAAL9529 [Arabidopsis thaliana], GRAB2 protein
GB:CAA09372 [Triticum sp.]"
/complement (join(<8331. .8855,9526. .9800,9933. .>10137))
/genes="F14G6.2"
/complement (join(8331. .8855,9526. .9800,9933. .10137))
/genes="F14G6.2"
/codon_start=1
/product="unknown protein; 10137-8331"
/protein_id="AAG51953.1"
/db_xref="GI:12323982"
/translation="MMLAVEDVLSLAGEERNELGPPGFRFHTDEELITFVLASKI
FHGSLGSIHISEVDLNRCEPMELPENAKMGEREMFYSLDRKYPTGLTNRAATAGY
WKATGSKVEFGGGQLVGMKTLVYKGRAPRGLTKWVHEVRLNDHSHRHTCK
EWWICRVNKTGDRKNGLIHNQISLHNHLSLSTHHHHEALPLLIPSNNKLTNF
PSLLYDDPHQNTNNNNFHLHGSGHNIDELKALINPVVSQLNIIFFSGNNNDEDDFD
FNLVQTEQSSNGNEIDVRDYLENPLFQEASVYLLGFSSSPGLHMLLDSPCPLGFL
"
/complement (15417. .18176)
/genes="F14G6.3"
/notes="similar to phosphate transporter GB:BAAL20522
[Catharanthus roseus]; contains Pfam profile: PF000083
Sugar (and other) transporter"
/complement (join(<15417. .16305,17467. .>18176))
/genes="F14G6.3"
/complement (join(15417. .16305,17467. .18176))
/genes="F14G6.3"
/codon_start=1
/product="putative phosphate transporter; 18176-15417"
/protein_id="AAG51941.1"
/db_xref="GI:12323970"
/translation="MPELSLSALDAARTQVHFKAIIIVAGMGLFTDAYDLFCIAPIM
KMSQIYIHKDSIGTALLSTSYAIALGLTALQGLIFGYLGDVRGRKRYVGLSLLIMVF
SFGCGFSYCTRRSCVMVSLGFLFVGLGIGDGYPLSATIMSEFANKRTRGAFIAA
VFSGGLGLMSAVTMVCLAPKNAGEGSEKTNVAGLETAPLAPESDIARWLLIMIG
ALPAALTFFYWRMLPMTARTVATLVENVVQAADQMORVMSVSMISOITEDSSSELBP
PSSSVKLFRRPLSLHGRDLFAASANWELVDVFTVSNLLLSOIFENFSKNPLNSNV
YDSAFVAKLAALVAACSTIPGWFTVYFIDIKLRVKIOMMGFFLMAVYLVAIGIPYS
WWSRHEKNTGFMVLYGLIIFPSNFGPNMTTIIIPAEFLPARFRSTCHGISGAOKF
GAIVGVFLWATRRHEEDGPDVQRVIAFLILGVGVCIAGMIVTYLFTRETWGRSLE
ENEDEIVTSAGSSPANELLRRQY"
/complement (18673. .19301)
/genes="F14G6.4"
/notes="predicted by genscan+"
/complement (join(<18673. .18926,19124. .>19301))
/genes="F14G6.4"
/complement (join(18673. .18926,19124. .19301))
/genes="F14G6.4"
/codon_start=1
/product="hypothetical protein; 19301-18673"
/protein_id="AAG51943.1"
/db_xref="GI:12323972"
/translation="MNAENNQTTHTSKIVSHVFCGTAKLGSVGPPIGLVDIGVSEV
AVIFRVLPGIEKNQDKICEIQREGRVCIQVIPAIPSDTGCLYRQVQVQLCPPG
PFSITNLPGQVDPRLFPSNFRSDGIFEVVVKLGVRIFTS"
/20920. .22411
/genes="F14G6.5"
/join(<20920. .20986,21073. .21148,21227. .21330,21407. .21472,
21739. .21821,21911. .21986,22064. .22182,22264. .22308,
22385. .>22411)
/genes="F14G6.5"
/join(20920. .20986,21073. .21148,21227. .21330,21407. .21472,
21739. .21821,21911. .21986,22064. .22182,22264. .22308,
22385. .22411)
/genes="F14G6.5"
/codon_start=1
/product="unknown protein; 20920-22411"
```

```
gene
/protein_id="AAG51945.1"
/db_xref="GI:12323974"
/translation="MAAISPLUSSQSFNSPRVTTIDRRCSISIAAISVLDSNBEQ
HRISRDHVMKRRDVMQIASSVFLPAISPAFAETNASAFRYVTTETNKFETSI
VQSTAITGLGPDFTRMESFGKEAFAETLVSGLDRSQKQVGVTAKLIDSRASKGFY
PYEYTLQNPGEARKHLYSAIGMATNGWYRLYTTVGTQFTDEESAQSSKIQTVKFSR
FI"
/complement (23340. .24808)
/genes="F14G6.6"
/notes="contains Pfam profile: PF00076 RNA recognition
motif. (a.k.a. RRM, RBD, or RNP domain)"
/complement (join(<23340. .23431,23548. .23785,23879. .23919,
24008. .24124,24234. .24378,24602. .>24808))
/genes="F14G6.6"
/complement (join(23340. .23431,23548. .23785,23879. .23919,
24008. .24124,24234. .24378,24602. .24808))
/genes="F14G6.6"
/codon_start=1
/product="putative RNA-binding protein; 24808-23340"
/protein_id="AAG51948.1"
/db_xref="GI:12323977"
/translation="MAYQVPQSGPHYLSNPFQDTTFTKVFVGLAWETQSETLRQH
EQYGTILEAVIADKNTGSKGYGFVTRDPAARRACADPTPIIDGRANCNLASIG
RRPPLPVAVINPMDLLPRTLQVLEVRHLEIIRISNHPHITTSKELSIIMVTAY
GPEYMSQSGLYSPVMOQYLQVYQVGAVNSPVYQQLSQTIPNGHGYTAVQGSY
VFGSHILQGGPTVSTMTTSSMPALQAPPSGIPGPVPVQSHIIVHSPQFMQSTASDQ
TTR"
/complement (26257. .27736)
/genes="F14G6.7"
/notes="similar to cinnamoyl-CoA reductase GB:CAA56103
[Bucalyptus gunnii]; contains non-consensus GG acceptor
splice site at exon 4"
/complement (join(<26257. .26401,26478. .26667,26773. .26935,
27062. .27247,27332. .27486,27593. .>27736))
/genes="F14G6.7"
/notes="contains nonconsensus splice site"
/complement (join(26257. .26401,26478. .26667,26773. .26935,
27062. .27247,27332. .27486,27593. .27707))
/genes="F14G6.7"
/notes="contains nonconsensus splice site"
/codon_start=1
/product="putative cinnamoyl-CoA reductase; 27707-26257"
/protein_id="AAG51951.1"
/db_xref="GI:12323980"
/translation="MAVKOKVCTGTAGGFIASWLVKFLSRGYTHGVTRDPCDEKND
HUKLDNASKNLKLFKADLFDDEGLFSAIDGCGVPHIASVPVFEHEELIKPALGTGK
NVLEACTETKQVQVWVSSIAAVVNPWPQVAKDECDWSQTYLHSLERYVYLAAT
LTERBALEWSKRNFDVVTLCPSVILGPRLQSLNSSLGLLKFIKGKIKSLSDSLY
LVDVRDADALLLVENREATGRYICNSHSLYTDLSLMEKLNKMPKRNPPESPTEVKE"
Query Match 61.7%; Score 593.8; DB 8; Length 100806;
Best Local Similarity 94.7%; Pred. No. 6.3e-83;
Matches 641; Conservative 0; Mismatches 22; Indels 14; Gaps 2;
QY 1 TGTGTTTCTCAGAAATAGCAGAAATATTTATAAAGCATGCAATCTCTTATAGATCG 60
Db 39933 TGTGTTTTTCAGAAATAGCAGAAATATTTATAAAGCATGCAATCTCTTATAGATTG 39874
QY 61 CGAAGTTTAAAAAACAATAGAAATGTTTACAATATTAACATGCGTGTGTTTATTCGATAACA 120
Db 39873 CGAAATTTAAAAA-----TTACAATATTACATGGGTTTTTATTTCGATAACA 39827
QY 121 TGACAAATATTTATTTATTTTCATGAGTTTTTATTTGATAGCATGACAAATATTAATATAT 180
Db 39826 TGACAAATATTTATTTATTTTCATGAGTTTTTGTGATAGCATGACAAATATTAATATAT 39767
QY 181 CAGTGTTAATACATGTTTTTGTCTTAAATACATGCATTTTAAATATCGACATTTGTTTT 240
Db 39766 TAGTGTTAATACATGTTTGTCTTAAATACATGCATTTTAAATATCGATATTTGTTTT 39707
QY 241 TAAATCAAATCTAATCTCTTATATATCAACGACATTCAGCGGAAATTCAGGTAAAGA 300
Db 39706 TAAATCAAATTTAATCTCTTATATCAACGACATTCAGCGGAAATTCAGGTAAAGA 39647
```

QY	301	GAATAATAAGAAATGAGAGATAGAGAGATTTCTATGGAAAAAGAAAGAGAAACATGTAGG	360	gene	or analysis and annotation"
Db	39646	GAATAATAAGAAATGAGAGATAGAGAGATTTCTATGGAAAAAGAAAGAGAAACATGTGGG	39587	gene	22865. .24293
QY	361	TGAAC-AAAATAAGAGATATGATATATTTATGAGAGGTGGTGAAGATTATTTTAG	419	CDS	/gene="AT4g08700"
Db	39586	TGAACAAAATAAAGAGATATGATATATTTATGAGAGGTGGTGAAGATTATTTTAG	39527		complement(join(22865. .23096,23227. .24293))
QY	420	GAGAGGAGAGAGAAATAGAAAAAGAAAAATGACATGCTGAATCTGAAAGAGATGAATTGT	479		/gene="AT4g08700"
Db	39526	GAGAGGAGAGAGAAATAGAAAAAGAAAAATGACATGCTGAATCTGAAAGAGATGAATTGT	39467		/notes="similarity to various hypothetical proteins of Arabidopsis thaliana
QY	480	GTTAAAGATCAAGAGAGAAAGAGAAAGCTCCATGGCTAAAGTCTCGTAAAGAGATGAAAA	539		Contains Prokaryotic membrane lipoprotein lipid attachment site AA82-92; Prokaryotic membrane lipoprotein lipid attachment site AA225-235
Db	39466	GTTAAAGATCAAAAGAGAGAAAGAGAAAGCTCTATGGCTAAAGTCTCGTAAAGAGATGAAAA	39407		contains EST gb:AI997220.1"
QY	540	GAACAAAGAGAGAGAAAGAGAAAGAGAAAGCTAAATAGACTAATTTGCCAAAATTT	599		/codon_start=1
Db	39406	GAATAAAGAGAGAGAAAGAGAAAGAGAAAGCTAAATAGACTAATTTGCCAAAATTT	39347		/product="putative protein"
QY	600	CTGTAGCCGACAAATACTATTTCGTCCAGGTTATTTTGTGTTATCTTTTGAAGTCAAAA	659	exon	/protein_id="CAB82106.1"
Db	39346	CTGTACCGACAAGTACTATTTCGTCCAGGTTATTTTGTGTTATCTTTTGAAGTCAAAA	39287		/db_xref="GI:7321059"
QY	660	GTTATTTCTTACATATA	676	intron	/db_xref="UniProt/TREMBL:Q9LD71"
Db	39286	GTTATTTCTTACATATA	39270		/translation="MLLKEDEGTRTSVPTQLMKLRTHMILVIFISIFLISAQA
RESULT 7	ATT32A17/c	104386 bp	DNA	linear	IAVLLGRFYNEGSKWISTLVQTCGFPIYLPLCLFPASHSSSSSCSEKTLVWIVL
LOCUS	Arabidopsis thaliana DNA chromosome 4, BAC clone T32A17 (BSSA project).				SLGFAIGLONLLYSPGLLYLSASTYSILGSSQLAFNGVFSYINSOKITCLILPSVLF
DEFINITION	Arabidopsis thaliana				LSVAVLSLDDSDNSPSGDSKWSYILGCLCTVFASLIYSLQSLAQFSENVLKST
ACCESSION	AL161813 AC007635				FSNVLMQIYTSLSVASCVAVIGLFPASGEWMLLSVEMBEHEGQIVYVLTLVGTAVSQ
VERSION	AL161813.1 GI:7321058				LGSVAVALIFLVSSLSFNLTGLSVITPLAAIAVFRHDKLITVKMAMLIAMFGFV
KEYWORDS	Arabidopsis thaliana (thale cress)				YIQNYLDLKVQRAEAQAEYLFLRAKVLRIYIRNRSTYEVKNGFTLENLGDTHV
SOURCE	Arabidopsis thaliana				VSDHDVADLETNMVESNKHVDDSTLTKASSSKRWIKGF"
ORGANISM	Arabidopsis thaliana				/gene="AT4g08700"
REFERENCE	1 (bases 1 to 96453)				/number=1
AUTHORS	McCombie,R.W., Spiegel,L.A., Huang,E.N., Nascimento,L.U., de la Bastide,M., Vil,D.M., Preston,R.R., Matero,A., Shah,R., O'Shaughnessy,A., Rodriguez,M., Shekher,M., Schutz,K., See,L.H., Swaby,I., Habermann,K., Dedhia,N.N., Mewes,H.W., Lemcke,K. and Mayer,K.F.X.				complement(23097. .23226)
JOURNAL	Unpublished				/gene="AT4g08700"
REFERENCE	2 (bases 95732 to 104386)				/number=1
AUTHORS	McCombie,R.W., Robben,J., Grymonprez,B., Bastiaens,I., Volckaert,G., Mewes,H.W., Lemcke,K. and Mayer,K.F.X.				complement(23227. .24293)
JOURNAL	Unpublished				/gene="AT4g08700"
REFERENCE	3 (bases 1 to 104386)				/number=2
AUTHORS	EU Arabidopsis sequencing, project.				/gene="AT4g08710"
TITLE	Direct Submission				join(24970. .25451,25915. .26692,26786. .26947,27034. .27477,27519. .27800)
JOURNAL	Submitted (10-WAR-2000) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: lemcke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de				/gene="AT4g08710"
COMMENT	On Mar 23, 2001 this sequence version replaced gi:5103797. Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.				/db_xref="GI:7321060"
FEATURES	Location/Qualifiers				/codon_start=1
source	1. .104386				/product="putative protein"
	/organism="Arabidopsis thaliana"				/protein_id="CAB82107.1"
	/mol_type="genomic DNA"				/db_xref="GI:7321060"
	/variety="Columbia"				/translation="MPPLKKRKGVTGTVKSTTEETATATKETAATKETAATKETA
	/db_xref="taxon:3702"				PTITKETAPTATKETAATKETAATKETAATKETAATKETAATKETAATKETAATKETA
	/chromosome="4"				EGEE
misc_feature	1. .22908				CNNNECLQEIDEETEAAPPPLSMYFPPEYVKKIKLSTRCYIHELLTTFDKLPPPSK
	/note="overlap to BAC T3F12; please refer to EMBL:AC002983				SERLWFENHPFQHIHPRDPRNHLRMLGMMMLLRARTARIEKKEAMFVINGVPIRYGI

```

exon      26786..26947
/gene="AT4g08710"
/number=3
intron    26948..27033
/gene="AT4g08710"
/number=3
exon      27034..27477
/gene="AT4g08710"
/number=4
intron    27478..27518
/gene="AT4g08710"
/number=4
exon      27519..27800
/gene="AT4g08710"
/number=5
gene       28303..31953
/gene="AT4g08720"
complement(join(28303..28494,28565..28761,29626..30162,
30333..31953))
/gene="AT4g08720"
complement(join(28303..28494,28565..28761,29626..30162,
30333..31953))
/gene="AT4g08720"
/notes="similarity to hypothetical protein F26B6.15 -
Arabidopsis thaliana, PID:g3242713"
/codon_start=1
/product="putative protein"
/protein_id="CAB82108.1"
/db_xref="GI:7321061"
/db_xref="UniProt/TREMBL:Q9LDY3"
/translation="MVVEKVIIVPRKDRAYSLMMNGKITLSMLESPICSKLGLDESKV
KLOKNYANVLGLTQSIHSDDEDVGVIYISSGEMWSRYPLTVDTVPPELPELSKVE
VYNKSGIRNGDGLSGNEDMDNATIVIGELDQGTLENGAIVVDDTDDEHIGSRVE
DRDDDDRGDEYVEPPDVESSQPKWEDGIGLTROEFPFSAALQEVVDYKGAFANSF
DVIYIKSDERYVLTCSKANCOWRIQASCVCQDTSIYSIRRYNKMHSCTRIKSSSTRMF
KRKGTPELVAALHPTDPGLLETTPPKVIMELVOTKLGVKVSYSTALRGKROAVCDLK
GAEDSYKDINCVLVYMLKKVNDGTVTYMKLDSGKFQYLFIALGASIEGFQAMRKVII
VQATHLKNYGGVVLFPASARDPNRHHYIIAVGLDGENDASGWFFPEKLLSVVPDTPPE
LVFMSDRNSLLIGIRNAYTAAHHGVCVWHLSQNVKAHSTNINRDVLAWRFMELSRIV
VTEFEPREYRVFKRYPTAVNYLEDSTVKEKWSVDNKFPVLVENVYLNHLWDAEKLK
VTELNTFRLEYNVYRGVEGNEYLVNLFHTKCSKVFDIQKYPGCIHALAAFIVPDSOTTR
TRGMELHELVSRYWAEALWALAYGTIYLVPGRSOWDIPODIKALKVLPLPRKKKGR
TKVLRTPSTGERPKRQRTONKERPRQSLQWLLRGTRTWDRLCESDPWVGEMKSAKRI
MQOLEVAMMEGIPPCPCGGRIILDIISEKDGKGRFYECTNRYKNDGLHIRKLWDKA
MVVEVNLREQVDNHHQKIQSUEYINQQMRSKLDEIQKXNGTLWLRICCGE"
complement(28303..28494)
/gene="AT4g08720"
/number=1
complement(28495..28564)
/gene="AT4g08720"
/number=1
complement(28565..28761)
/gene="AT4g08720"
/number=2
complement(28762..29625)
/gene="AT4g08720"
/number=2
complement(29626..30162)
/gene="AT4g08720"
/number=3
complement(30163..30332)
/gene="AT4g08720"
/number=3
complement(30333..31953)
/gene="AT4g08720"
/number=4
34618..36007
/gene="AT4g08730"
complement(join(34618..34852,35539..35690,35828..36007))
/gene="AT4g08730"
complement(join(34618..34852,35539..35690,35828..36007))
/gene="AT4g08730"
/codon_start=1
```

```

/product="hypothetical protein"
/protein_id="CAB82109.1"
/db_xref="GI:7321062"
/db_xref="UniProt/TREMBL:Q9LDC3"
/translation="MEGFTNFLOQVMEKTKVDLLARKVLDLTSKMDLCATNVNLLTSKWN
LCATNSDLKAMLTSMKGNVPAKTKLWMAASDSEDSSDDIELCFRFLIKIVIQ
LRPSYFIILIPAYIDILKEGNKALELNGSDMAGKELVVKLTALLMRDSYLGYSNCGGIGG
RFGGRPGSVGGIFGGRCGGLGQYYAQV"
complement(34618..34852)
/gene="AT4g08730"
/number=1
complement(34853..35538)
/gene="AT4g08730"

Query Match      60.2%; Score 579.2; DB 8; Length 104386;
Best Local Similarity 90.9%; Pred. No. 1.1e-80;
Matches 651; Conservative 0; Mismatches 58; Indels 7; Gaps 3;

QY      1  TGTGTTTCTCAGAAATAGCACGAAATATTATATAAAAGCATGCAATCTCTTATAGATCG 60
Db      87837 TGTGTTTCTCTGAATAGCACCAGATATTATATAAAAGCATGCAATCTCTTATAATCG 87778

QY      61  CGAAGT-TTAAAAAACATATAGAAATTGTTACAAATATTACATGGGTTTTTATTGGATAAC 119
Db      87777 CGAAGTGTTTAGAAACATATAGAAATTGTTACAAATTTACATCGGTTTTTATTGGATAAC 87718

QY      120 ATGACAAATATTATTATTTTTCATGAGTTTTTATTGGATAGCATGACAAATATAATATA 179
Db      87717 ATGACTAATATTATATATTTTCATGAGTTTTTGTTCGATAGCATGACAAATATAATATA 87658

QY      180 TCAGTGTTAATACATGTTTTTGTTCCTTAAATACATGCATTTTAAAAATCAGACATTTGTT 239
Db      87657 TCAGTGTTAATAACATGTTTTTGTTCCTTAAATACATGCATTTTAAAAATCAGACATTTGTT 87598

QY      240 TTAATAATCAAACTCTAATCTCTTATATACAAACGACATTCACGCGAAAAATTCAGGTAAAAAG 299
Db      87597 TTAACATCAAACTAATCTCTTATATACAAACGACATTCATGGAGAAATTCAGGCAAAAAG 87538

QY      300 AGAAAAATAAGAAATAGAGATAGAGAGATTCTTATCGAAAAAGAAAGAGAGAACATGTAG 359
Db      87537 AGAAAAATAAGAAATAGAGATAGAGAGATTCTTATCGAAAAATAAAGAGAGAACATGTGG 87478

QY      360 GTGACAAAAATAAGAGATATGATGATATATTTTATGAGAGGTGGTGAAGATTATTT--- 416
Db      87477 GTGAACAAAAATAAGAGATATGATGATATATTTTATGAGAGGTGGTGAAGATATTTCTAGG 87418

QY      417 --TAGGAGAGGGAGAGAGAAATAGAAAAAGAAAAATGACATGGTGAATCTGAAGAAAGATGA 474
Db      87417 AGAGGGAGAGAGAAATAGAAATAGAAAAAGAAAAATGACATGGTGAATCTGAAGAAAGATGA 87358

QY      475 ATTGTGTTAAAGATGAAGAGAGAAAGAACTCCATGGCTTAAAGTCTCGTAAAGAGAGATG 534
Db      87357 ATTGTGTTAAAGATGAAGAGAGAGAAAGAACTCCATAGCTAAAGTCTCGTAAAGAACATG 87298

QY      535 AAAAAAGAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 594
Db      87297 AAAAAAGAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 87239

QY      595 AATTTCTGTAGCGCGCAAAATACTATTGTGTCGAAGTTATTTTGTGTTATTTCTTTTCAAGT 654
Db      87238 AATTTCTGTAGTACCAAGTACTATTGTTCCAGGTTATTTGTTTATTCTTTTCAAGT 87179

QY      655 CAAAAGTTATTTCTTACATATATCTCTAAAAATATAGCCGATACCAATTTTTCACA 710
Db      87178 CAAAGGTTATTTGTATATATCTCTTTTATAAAAAAAATAATAACGTTTGCAA 87123
```

```

RESULT 8
ATCHRIV25/c
LOCUS      ATCHRIV25      179771 bp      DNA      linear      PLN 16-MAR-2000
DEFINITION Arabidopsis thaliana DNA chromosome 4, contig fragment No. 25.
ACCESSION  AL161513
VERSION     AL161513.2      GI:7267524
```

KEYWORDS

SOURCE

ORGANISM

Arabidopsis thaliana (thale cress)
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi
 1 (bases 1 to 26363; 134798 to 179771)
 Spiegel, L.A., Huang, E.N., Nascimento, L.U., de la Bastide, M.,
 Vill, D.M., Preston, R.R., Matero, A., Shah, R., O'Shaughnessy, A.,
 Rodriguez, M., Shekher, M., Schutz, K., See, L.H., Swaby, I.,
 Habermann, K., Dedhia, N.N., Mewes, H.W., Lemcke, K. and Mayer, K.F.X.
 Unpublished

JOURNAL

REFERENCE

AUTHORS

2 (bases 25642 to 34296)
 Robben, J., Grymonprez, B., Bastiaens, I., Volckaert, G., Mewes, H.W.,
 Lemcke, K. and Mayer, K.F.X.

JOURNAL

REFERENCE

AUTHORS

3 (bases 33706 to 33707)
 Lamar, B., Stoneking, I., Stumpf, J., Mewes, H.W., Lemcke, K. and
 Mayer, K.F.X.
 Unpublished

JOURNAL

REFERENCE

AUTHORS

4 (bases 1 to 179771)
 EU Arabidopsis sequencing project.

TITLE

JOURNAL

Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
 Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
 lemcke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de Project
 Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
 Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
 E-mail: michael.bevan@bbsrc.ac.uk

COMMENT

Information on performance of analysis and a more detailed
 annotation of this entry and other sequences of chromosomes 3, 4
 and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>
 this fragment has an overlap with ATCHRIV24 at the 5' end and an
 overlap with ATCHRIV26 at the 3' end.

FEATURES

source

1. 179771
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /variety="Columbia"
 /db_xref="taxon:3702"
 /chromosome="4"
 8902..10347
 /gene="AT4g08820"
 join(8902..9093,9193..9553,9686..10347)
 /gene="AT4g08820"
 /note="similarity to various hypothetical proteins of
 Arabidopsis thaliana Chromosome II, III and IV"
 /codon_start=1
 /product="putative protein"
 /protein_id="CAB78007.1"
 /db_xref="GI:7267525"
 /db_xref="Uniprot/TREMBL:Q9LE84"
 /translation="MSVSNKGSVGGILKPEENVIDAFVRRVRLFPNGDEGPVIT
 IGBEVLAMNGLKRCMIVRFIRFELEEYLAALTGPMRVFGNYLMVDWSNPF
 PLRDDVITVTNLTNPVNYHRCLEELIARGLKLKVDLNTITFGGRFARVCI
 EVNLAKPLKTVLINDRYFVAYEEVEEGFTVRQSGIRVEKPAQKMFPAAGSGGRS
 NRSRLRLPKNGVEISNRFGGLEDMVSAEIGVAILEGSKENKYGKGNRKESGVT
 QVKEAQNVTFKKGKGNKGPWKNGGLIIMEPNGPRQKQSVNKPVRGLIFGPTK
 GESSNLVDALSANGKRLRVEQRSGVRPGGAFVSGLETRSGEMDMSELPQNNNIQQ
 SKELLCDTAIVV"
 8902..9093
 /gene="AT4g08820"
 /number=1
 9094..9192
 /gene="AT4g08820"
 /number=1
 9193..9553
 /gene="AT4g08820"
 /number=2
 9554..9685
 /gene="AT4g08820"
 /number=2
 9686..10347

gene

CDS

/gene="AT4g08820"
 /number=3
 10891..11438
 /gene="AT4g08830"
 join(10891..11025,11244..11393,11430..12120,12361..13089,
 13162..13585,13634..14348)
 /gene="AT4g08830"
 /note="similarity to various putative reverse
 transcripts of Arabidopsis thaliana
 Contains Aminoacyl-transfer RNA synthetases class-II
 signatures AA699-708"
 /codon_start=1
 /product="putative protein"
 /protein_id="CAB78008.1"
 /db_xref="GI:7267526"
 /db_xref="GOA:Q9LDL8"
 /db_xref="Uniprot/TREMBL:Q9LDL8"
 /translation="MSPDLSLEFGEWINRVSLIDMGFSGNKYTKWGRVNTFVAKRLD
 RLKKNRDVFQDVKKKEELMREIQVQDALDVHQSDMLKKEEDLIKAFDKSREKWI
 ALGDRNTNYFHTTIIIRRNQRVERLHNEGTWIFDAKELEALAVQYQLYSMDVVL
 OVVPPLPDGLERLTREEVLSLNKPLATEVEAIRSMGKYKAPGPDGYQPIFYQSCW
 EVVGNVIRFALDPFSSGILPQTNDAVLVLPKPKPERMNOFRPLSLCNAIFKQMIT
 KMVLRKLVIGKLIQPSQSFIPGRSLDNIIVVQAVHSLNKKRKRCLCHMIDR
 AVAKEMKSGISGGQPKISHICFADDLILFAEASVQIRVIRILLETFCIAGSQKVS
 LDKSKIFFSKNVSRLDKLTSKESGISTRELKYLGMPTLQRIINRDKTDFEVLKVS
 SRLAKGKRSLSFAGRLTLTKSVLSLPIHTMTSTISLPQSTLEGLDKLARVLLGSSA
 EKKLLHLAWDRVCLPKSGGLGIRTSKCMNKALSVKGMWRLINDRYSLKARILRSKY
 RVGLREVSRGSRVVGNGRDILFWSDNLSHEALINRAVIEIPNSEKEKLRVKDLWAN
 HGDRLDKIEPIYSIHTRLLEAAVVDVSVTGARDLSWGSYADGVFTVKSAYRLITSD
 HDPNPNAAPFDRLRVVALERVKTLFHHIGDTSVCQCKGGDETILHLVKDCPSIAG
 IWRLLVQVQSRYSDFNGSLFGMLVNLGMKNAGTYAWATLFAIVVMWSKRCGYVF
 GEVKKCRDRKFFRDAAAEVSHAHAIHSONGLRTRVERLVANPKPDGEMWKLNTDGA
 SRNLGLATTGGVLRDGIHWCQGFALDIGVCSAPLAELGWVYVYGLYMAERRRTRVE
 LEVDSLVVGFLLTIGISDTHSLSLVRLCHGFL"
 10891..11025
 /gene="AT4g08830"
 /number=1
 11026..11243
 /gene="AT4g08830"
 /number=1
 11244..11393
 /gene="AT4g08830"
 /number=2
 11394..11429
 /gene="AT4g08830"
 /number=2
 11430..12120
 /gene="AT4g08830"
 /number=3
 12121..12360
 /gene="AT4g08830"
 /number=3
 12361..13089
 /gene="AT4g08830"
 /number=4
 13090..13161
 /gene="AT4g08830"
 /number=4
 13162..13585
 /gene="AT4g08830"
 /number=5
 13586..13633
 /gene="AT4g08830"
 /number=5
 13634..14348
 /gene="AT4g08830"
 /number=6
 16002..16053
 18436..20750
 /gene="AT4g08840"
 join(18436..19424,19870..20032,20124..20287,20492..20750)
 /gene="AT4g08840"
 /note="similarity to probable regulatory protein YGL014w

repeat_region

gene

intron

exon

-Saccharomyces cerevisiae, PIR2:S64016	
Contains Prokaryotic membrane lipoprotein lipid attachment	
Site AA82-92"	
/codon_start=1	
/product="putative protein"	
/protein_id="CAB78009.1"	
/db_xref="GI:7267527"	
/db_xref="GOA:Q9LDW3"	
/db_xref="UniProt/trEMBL:Q9LDW3"	
/translation="MDFGFFGDLRQGSFTDLGFNGFPLTSVNSNGHFSGDRTNTP	
FNLRLKDTTSLMADGDMGLCNLSKMSISDRSNFFNHSPFSGYCYQGRESFPHG	
EASSMRGFGVGDVHRFQDLRVRASPHGESAMSSVYGDGDYHRLRFLALQEAQNP	
NPRCTFNMSLLNRDYMLELHFNQOIREDFSI.VPKSPLAFHEERIL.PPFSAMGSR	
ELDGSACKMKNEKEDSLASMDVDSYGVYLMMAKQDLCRLLQKFVDEGNFVDMILFK	
EVINNVIEGTDFPGNYLLOKLEVCNEEQRTQILRLRLTSKPGLLVVKISINNYGTRV	
QKLIETATTKQISLVKLSALVPGLFSLFRELNGNHVILNCLKFFSPDNKFLIEAATK	
FCIEATTHGGCVLQRCVSYSGEHEKLVDIEISRNLSLLAQDPGNLYLVQVIEKK	
VGVNVLFELRGNYVKLATQKFGSHVVEKLCRYYPESRSQIVNELVSLNFGVLLQDP	
YANYVIQCALSRTKVS"	
exon	18436..19424
/gene="At4g08840"	
/number=1	
intron	19425..19869
/gene="At4g08840"	
/number=1	
exon	19870..20032
/gene="At4g08840"	
/number=2	
intron	20033..20123
/gene="At4g08840"	
/number=2	
exon	20124..20287
/gene="At4g08840"	
/number=3	
intron	20288..20491
/gene="At4g08840"	
/number=3	
exon	20492..20750
/gene="At4g08840"	
/number=4	
gene	23831..27580
/gene="At4g08850"	
gene	complement(join(23831..24120,24787..27580))
/gene="At4g08850"	
CDS	complement(join(23831..24120,24787..27580))
/gene="At4g08850"	
/notes="Est GB:F14413 is reverse and marks the end of the	
gene	
similarity to receptor protein kinase-like protein -	
Arabidopsis thaliana, PIR2:T05898	
Query Match 60.2%; Score 579.2; DB 8; Length 179771;	
Best Local Similarity 90.9%; Pred. No. 9.8e-81;	
Matches 661; Conservative 0; Mismatches 58; Indels 7; Gaps 3;	
QY	1 TGTGTTTCTCAGAAATAGCAGCAATATTATATAAAGCATGCAATCTCTTATATATCG 60
Db	17747 TGTGTTTCTCTGAAATAGCAGCATATATTTATATAAAGCATGCAATCTCTTATAATCG 17688
QY	61 CGAAGT-TTAAAAAACAATAGCAATTTGTTACAAATTTACATGGGTTTTTATTGGATAAC 119
Db	17687 CGAAGTGTTTAAGAAACATATAGAAATGTTTACAAATTTACATCGGTTTTTATTGGATAAC 17628
QY	120 ATGACAAATATTATTATTTCATGAGTTTTTTATTGGATAGCATGACAAATATAATATA 179
Db	17627 ATGACTAATATTATATATTTTCATGAGTTTTTTGTTGGATAGCATGACAAATATAATATA 17568
QY	180 TCAGTGTTAATACATGTTTTGTTCTTAAATACATGCAATTTTAAATCAGCAATTTGTT 239
Db	17567 TCAGTGTTAATACATGTTTTGTTCTTAAATACATGCAATTTTAAATCAGCAATTTGTT 17508
QY	240 TTAAAAATCAAAATCTAATCTCTTATATCACAAACGACATTTGACGGAATTCAGGTAAAAAG 299

Db	17507	TTATCATCAATCTAATCTCTTTATATCACACGACATTGATCGAGAAATTCAGGCAAAAAG	17448
QY	300	AGAAATAAAGAAATGAGAGATAGAGATTTCTATGAAAAAAGAAAGAGAGAACATGTAG	359
Db	17447	AGAAACCAAGAAATGAGAGATAGAGATTTCTATGAAAAAATAAGAGGGAACATGTGG	17388
QY	360	GTCACAAATAAAGAGATATGATATATTTTATGAGAGCTGCTGAAGATTATTT---	416
Db	17387	GTGAACAAATAAAGAGATATGATATATTTTATGAGAGCTGCTGAAGATTATTTAGG	17328
QY	417	--TAGGAGAGCGGAGAGAGAAAATAGAAAAAGAAAAATGCATGCTGAATCTGAAAGAACATGA	474
Db	17327	AGAGGAGAGAGAAATAGAAATAGAAAAAGAAAAATGCATGCTGAATCTGAAAAAGATGA	17268
QY	475	ATTGTTTAAAGATGAAGAGAGAAAGAACTCCATGGCTAAAGTCTCTGTAAGAAGATG	534
Db	17267	ATTGTTTAAAGATGAAGAGAGAAAGAACTCCATAGCTAAAGTCTCTGTAAGAAGATG	17208
QY	535	AAAAAGAAACAAAGAGAGAAAGAGAAAGAGAAAGCTAAAAATAGACTAATTTGCCAA	594
Db	17207	AAAAATAAATAAGAGAAAGAGAAAGAGAAAGAGCTAAAAATAGACTAATTTGCCAA	17149
QY	595	AATTTCTGTAGCCGACAAATACTATTTTGTCCAAAGTTATTTTGTGTTATTTCTTTTCAAGT	654
Db	17148	AATTTCTGTAGTAGACAGTACTATTTTGTCCAAAGTTATTTTGTGTTATTTCTTTTCAAGT	17089
QY	655	CAAAAGTTAATTTCTTACATATATCTCTTAAAAATATAGCCGATACCAATTTTCCACA	710
Db	17088	CAAGGTTATTTTGTATATATATCTCTTTTATAAAAAAATAATAACGTTTGCAAA	17033
RESULT 9			
AF071527		95519 bp DNA linear PLN 06-MAY-1999	
LOCUS		Arabidopsis thaliana BAC F9H3, from chromosome IV near 18.8 cm,	
DEFINITION		complete sequence.	
ACCESSION		AF071527	
VERSION		AF071527.1 GI:3236479	
KEYWORDS		HTG.	
SOURCE		Arabidopsis thaliana (thale cress)	
ORGANISM		Arabidopsis thaliana	
		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
		Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;	
		rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsiis.	
REFERENCE		1 (bases 1 to 95519)	
AUTHORS		Huang, E.N., Parnell, L.D., de la Bastide, M., Schutz, K.,	
		Habermann, K., Dedhia, N.N. and McCombie, W.R.	
		Genomic sequence of Arabidopsis Thaliana BAC F9H3, chromosome IV,	
		18.8 cm	
TITLE		Unpublished	
JOURNAL		2 (bases 1 to 95519)	
REFERENCE		Huang, E.N., Parnell, L.D., de la Bastide, M., Schutz, K.,	
AUTHORS		Habermann, K., Dedhia, N.N. and McCombie, W.R.	
		Direct Submission	
TITLE		Submitted (10-JUN-1998) Cold Spring Harbor Laboratories, Lita	
JOURNAL		Annenberg Hazen Genome Center, 1, Bunting Road, Cold Spring	
		Harbor, NY 11724, USA	
REFERENCE		3 (bases 1 to 95519)	
AUTHORS		Parnell, L.D. and McCombie, W.R.	
		Direct Submission	
TITLE		Submitted (31-DEC-1998) Lita Annenberg Hazen Genome Sequencing	
JOURNAL		Center, Cold Spring Harbor Laboratory, 1 Bunting Road, Cold Spring	
		Harbor, NY 11724, USA	
REMARK		Arabidopsis thaliana BAC F9H3 from chromosome 4, near 18.8 cm	
REFERENCE		4 (bases 1 to 95519)	
AUTHORS		Parnell, L.D. and McCombie, W.R.	
		Direct Submission	
TITLE		Submitted (01-FEB-1999) Lita Annenberg Hazen Genome Sequencing	
JOURNAL		Center, Cold Spring Harbor Laboratory, 1 Bunting Road, Cold Spring	
		Harbor, NY 11724, USA	
REMARK		Arabidopsis thaliana BAC F9H3 from chromosome IV near 18.5 cm	
REFERENCE		5 (bases 1 to 95519)	
AUTHORS		Parnell, L.D.	

TITLE
JOURNAL

REMARK
COMMENT

Direct Submission
Submitted (06-MAY-1999) Lita Annenberg Hazen Genome Sequencing Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724
Arabidopsis thaliana BAC F9H3 from chromosome IV near 18.8 cM
BAC F9H3 is assigned between YACs C1C47 and C1C8B1 and maps to near 18.8 cM on the Lister & Dean RI map. Position 1 of F9H3 is oriented toward the centromere and position 95519 is oriented toward the telomere. For more information on the mapping, sequencing and annotation of F9H3, please see <http://www.cshl.org/arabweb/F9H3-titlepage.html>. A graphic view of our annotation is also available at this url. Gene models are built with exons predicted by GenScan (<http://CCR-081.mit.edu/GENSCAN.html>), MZEF (<http://www.cshl.org/genefinder>) and GRAIL (<http://compbio.ornl.gov/tools/index.shtml>) and with splice sites predicted by NetPlantGene (<http://www.cbs.dtu.dk/netpgene/cbsnetgene.html>). Alternate exons not used in building the gene models are presented on the web pages associated with F9H3. Genes are numbered according to the scheme BAC.gene number. Typically, these numbers progress from 1 upwards as one moves from position 1 of the BAC. Protein sequences encoded by the genes are assigned to a functional category with the aid of similarity searches and comparison to the Prosite (<http://expasy.hcuge.ch/sprot/prosite.html>) and Pfam (<http://pfam.wustl.edu/>) libraries. A description of these categories can be found at <http://muntjac.mips.biochem.mpg.de/arabi/>. Genomic repeats are typically located by TBLASTX analysis and an attempt is made to classify the function of each repeat as either transposon, putative microsatellite, LINE, direct repeat, centromeric repeat, etc.

If you have any questions or confirmatory or contradictory evidence concerning the annotation of F9H3, please direct email to Larry Parnell at parnell@cshl.org.
BAC F9H3 was sequenced as part of the arabidopsis genome sequencing effort of the Cold Spring Harbor Consortium. For additional information, please see <http://www.cshl.org/arabweb>.
Fingerprint data indicate F9H3 overlaps with F4C21 toward the telomere and with T5L23 toward the centromere. F9H3 contains marker m1233 at 18.8 cM on the Lister & Dean RI map.

FEATURES
source
1..95519
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="IV"
/map="18.8 cM"
/clone="BAC F9H3"
/notes="hybridizing YAC is C1C8B1"
misc_feature
1..31783
/notes="overlap with T5L23, GenBank accession number AC005142, from position 26 to 31808; Region: overlap with T5L23"
851..1719
/gene="F9H3.20"
/pseudo
/evidence=not experimental
complement (981..1120)
/notes="functions=unclassified; similar to F6N23, GenBank accession number AF058919"
rpt_type=dispersed
3173..3969
/notes="functions=putative transposon; similar to FCAJ1. GenBank accession number Z97336; moderately similar to reverse transcriptase gene of DM15 non-LTR retrotransposon, GenBank accession number L47193"
rpt_type=dispersed
5889..6303
/notes="functions=chloroplast_repeat"
rpt_type=dispersed

gene

complement (6653..12091)
/gene="F9H3.19"
/notes="encodes putative calcium channel; genomic copy of EST W43766; genomic copy of EST T43251; EST T43251 apparently encodes intronic sequence; possible alternate exon 1 from position 12772 to 12846 is predicted by a NetPlantGene splice donor but has no EST coverage; gene model last edited on 27 Jan 99"
/evidence=not experimental
complement (join(6653..6746,6833..6870,6991..7062,7175..7261,7436..7526,7682..7746,7985..8153,8229..8292,8380..8487,8757..8814,8891..9034,9127..9192,9311..9383,9530..9591,9745..9822,9929..10034,10127..10293,10391..10451,10557..10580,10644..10774,10890..10930,11292..11403,11500..11643,11972..12091))
/gene="F9H3.19"
/note="functional catalog ID=07.01"
/codon_start=1
/evidence=not experimental
/product="putative calcium channel"
/protein_id="AAD11598.1"
/db_xref="GI:4206210"
/translation="MEDPLIGRSLGGGTDRVRSEATHGTPTFQKAALVDLAEDGIGLPVBIQSSSGESARIYFTRDLIWSLNFALLFTFQKDNVNRQEHVHLAARDYYLIGELPYLTNAESIYEASSRIFWTSRLNLVKVACVILFVDVLVLYSLP AFDPLPFIAPYVRVIFILSIQRQCADHVMELRDLTVLLSGMLGYLNTLALWMLPLASMTAFVMEFTQGLTVFTSYGATLYQMFTLFTSNPDVMIIPAYKSSRWSVFF VLYLVGVFVNTLLIIVAVYDSFKEQAKQVSGMDQMKRMLEKAKGLIDSDKNGEID KQCILFQNLINIRTLKISKEEFGILFDELDDTRDKINKDEFADCCALRFPQK ECVSLFHFHPQIYHSALSQRAFVRSFNFGVIAFILINFIIVAVVETLIDIESS AOKPQVAFVFGWIRYLLMLKLIYTYGFENYWRGNFDFLTWTVVIGETATFIT PONTFPGNGEWIRYLLMLKLIYTYGFENYWRGNFDFLTWTVVIGETATFIT LCTYSGVQVFGGLVAGNKLFTLELAEDDYLLENFNDYNGMTVLFNLVAGNQV VMESYKDIATGTMWSITYVSFVYITLLILNLVAVLEAFETLELDEEBEKCQGD SEKRNRRSAGSKSRQVDTLLHMLGDELSPKCSSTDT"
complement (13149..18831)
/gene="F9H3.18"
/note="encodes putative glucan synthase component; gene model last edited on 5 Jan 99"
/evidence=not experimental
complement (join(13149..14597,14778..16800,16961..18831))
/gene="F9H3.18"
/notes="similar to 1,3-beta glucan synthase; functional catalog ID=01.05.99; functional catalog ID=09.01"
/codon_start=1
/evidence=not experimental
/product="putative glucan synthase component"
/protein_id="AAD11597.1"
/db_xref="GI:4206209"
/translation="MSLRHRTVPQTGRPLAAEAVGIEBEPYNIIPVNNLADHPSLR FEVRAAAALKTGDLRPPVQWPSHYDLDLMLALFTGFKDNVNRQEHVHLA NAQMRSLPPPDNIDSLDSAVRRFRKLLANYSWCYLGKKSNIWISDRNPSREL LVLGYLLWGEAANRFPPECIYIFHNWSELNKILEDCLENTGQPYLPUSGEN AFYLVGVVPIYDTIQAEIDESKNGTVAHCKWRYDDINEYFTWDFCSKLKWPDLGS NFFKSKGSGVGTGFRERTFFLYRSFRLWMLALFLQAAIIIVAEERKPDTSVTR QLNALKARDQVORLLTVFLTWSGMRLLOAVLDAASQYPLVSRTRKRPMLMKVIA AAVIIVATVLYTNIWKQKRODWSNAATKIYQFLYAVGAFLEILALALFIPW MRNLEETNWKIFPALTWNFQGSFVGRGLRGLVDNIKYSTFWIATKFTFSFPL QVKPMIKSKLLNLDKVDYEHQFYQSNRFSVALLPVLIIYDQIWIYAIYS IVGAVVGLDFHLGEIRLRFQFASAIQFNLMPEQLNARGFKNGKDGHR LKLYRGFRPFKKLESQVANKFALINWEIIAFLREEDIVSREVELLEPLKNSWDV TIVRWPCLLNCHELLALSOARELIDAPKWLWHKICKNEYRCACVAVDYSLKHLI SIUKVDTEHSIIIVFQIINOSIQSEQFTKFRVLLPKIYETLOKLVGLVNDIEDT SRVNVNLSQSLYEIATQFFIEKTKTEQSLNEGLTRDPASKLLFQNALRDPASNE D FYQVRRLHILTSRDSMHSVPVNLRRRIAFSSNLSFNMMPHAPQVEKQMSVLT PYSEEVYVSKQLRNEDGSLTYLQTIYADEWKNFKRMREGIKTGDSLWTK LRDRLWASRYGQTLARTVRGMWYVSTLYKHEYGTALEKMTFYVVAQCIYGSOKAKKE ELGQSDGFVSENDRSSLRASSSVTLKHEYGTALEKMTFYVVAQCIYGSOKAKKE PQAREILYLMKONEALRIYVDEVPAGGETDYSLVKYDHOLEKVEIFRVLKPGP VKLGECKPENQNHAMIFTRGDVAVQTIIDMNQDSYFEBALKRNLLQYNNHYHGRKPTI LGVREHIFTGSSVSLAWFMSAQETSFTLQGRVLANPLKVRMHYHPDVPDFRWFLSR GGISKASRVINISDIFAGFNCTLRGNVTHHEIYIQVGRGDVGLNQIISNFEAKVASG NGEQVLSRDVYRLGHRLDFFRMLSPFTYTTVTFGFFNTMMVILTVYAFWGRVYLAISGV

EKSLADSTDTNAAAGVILNQOFLIQLGFTALPMIVENSLBEGFLAIWFIWMQIQ LSAFTFMSGTRAHFGRITLHGGAKYRATGRGFVVEHKGFENTRLVARSHFVKAI EGLNLIIVYASHPSIAKSLDIYIAMITTSWFLVSIWMAFPFVFNPSGPFMLKTVYDFE DPMNTWYQCRISTKSEQSWERKWTYBEQDLHRLNTGKAGLFVSIILVLRFFQYQYIVY QKLIANGSTSLFVLPFSWYIYFAIFVLFLVIOVARDKYSAKAHIRVRLVFFQYQYIVY LVTVALLETSHFSFIDFTSLAFIPFGWILLIAOTQRKWLKNYTFIWNVUSVARM YDLIFGILLMVPAFLSWMRPGSQMQTRILFNEAFSRGLRIMQIVTGKSKGDV" complement (20861. .22302) /genes="F9H3.17" /notes="encodes hypothetical protein; gene model last edited on 5 Jan 99" /evidence=not experimental complement (join(20861. .21094,21937. .22039,22145. .22302)) /genes="F9H3.17"									
gene									
Query Match 59.4%; Score 571.6; DB 8; Length 95519; Best Local Similarity 92.2%; Pred. No. 1.7e-79; Matches 636; Conservative 0; Mismatches 49; Indels 5; Gaps 3;									
CDS									
QY	1	TGTGTTTCTCAGAAATAGACGCAAAATATTATATAAAGCATGCAATTCCTTATAGATCG	60						
Db	23202	TATGTTTCTCTGAATAACACGATATATTATTATAAAGCATGCAATTCCTTATAGATCG	23261						
QY	61	CGAAG-TTTAAAAACATATAGAATTGTTTACAATATTACATGGGTTTTTATTGGATAAC	119						
Db	23262	CGAAGTTTTTAAAAACATATAGATTGTTTACAATATTACATGAGTTTTTATTGGATAAC	23321						
QY	120	ATGACAAATATTATTATTTCATGAGTTTTTATTGGATAGCATGACAAATATTAATATA	179						
Db	23322	ATGACAAATATTATTATTATACATGAATTTTGTGATAGCATGACAAATATTAATATA	23381						
QY	180	TCAGTGTTAATACATGTTTGTCTTAAATACATGCATTTTAAATCAGACATTTGTT	239						
Db	23382	TCAGTGTTAATACATGTTTGTCTTAAATACATGCATTTTAAATCAGACATTTGTT	23441						
QY	240	TAAAAATCAATCTAATCTCTTATATACAAACACATTTGACGCAAAATTCAGTGAATAAG	299						
Db	23442	TAAAAATCAATATATATCTCTTATATACAAACACATTTGATGAAGATTCAGGCAATAAG	23501						
QY	300	AGAAAAATAAGAATAGAGATAGAGAGATTCTTATGGAAGAAAAAGAGAGACATGTAG	359						
Db	23502	AGAAAAATAAGAATAGAGATAGAGAGATTCTTATGGAAGAAAAAGAGAGACATGTGG	23561						
QY	360	GTGAACAAATAAAGAGATATGATATATTTTATGAGAGGTGGTGAAGATTAATTTAG	419						
Db	23562	GTGAACAAATAAAGAGATATGATATTAATTTTATGAGAGGTGGTGAAGA-TAATCTAG	23620						
QY	420	GAGAGGGAGAGAAATAGAAAAAGAAATGACATGCTGAATCTGAAGAAGATGAATTGT	479						
Db	23621	GAGAGGGAGAGAAAGTAGAAAAAGAAATGACATGCTGAATTTGAAGAAGATGAATTGT	23680						
QY	480	GTTTAAAGATGAAGAGAGAAGAGAAAGCGCTAAATAGACTAATTTGCCAAATTT	539						
Db	23681	GTTTAAAGATGAAGAGAGAAGAGAACTTCATGCTAAAGTTTCGTAAGAAGATGAATAA	23740						
QY	540	GAACAAAGAGAGAGAGAGAAAGAGAAAGCGCTAAATAGACTAATTTGCCAAATTT	599						
Db	23741	AAAAGAA---AAAGAGAGAAAAAGAGAAAAATATAATAGACTAATTTGCCAAATTT	23797						
QY	600	CTGTAGCGCAAAATACTATTTCGTCGAAGGTTATTTTGTGATCTTTTGAAGTCAAAA	659						
Db	23798	CTGTAGCAGACAAGTACTATTTCGTCGAAGGTTATTTTGTGATCTTTTGAAGTCAAAA	23857						
QY	660	GTTTATTTCCTACATATACTCTAAATAATA	689						
Db	23858	GTTTATTTCCTACATATACTCTATATATA	23887						

RESULT 10
AC005142/c
LOCUS
DEFINITION
Arabidopsis thaliana BAC T5L23 from chromosome IV, near 19 cm,
complete sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AC005142
AC005142.2
GI:4263038
HTG.
Arabidopsis thaliana (thale cress)

Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsiis.

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

1 (bases 1 to 116448)
Zhong,J., Ma,P., Parnell,L.D., Chen,C.-N. and Chen,E.Y.
BAC T5L23 from chromosome IV, position 19 cm
Unpublished
2 (bases 1 to 116448)
Zhong,J., Ma,P., Parnell,L.D., Chen,C.-N. and Chen,E.Y.
Direct Submission
Submitted (19-JUN-1998) Applied Biosystems Division of
Perkin-Elmer, 850 Lincoln Centre Dr., Foster City, CA 94404, USA
3 (bases 1 to 116448)
Parnell,L.D. and Chen,E.Y.
Direct Submission

REFERENCE
AUTHORS
TITLE
JOURNAL

On Feb 22, 1999 this sequence version replaced gi:3241935.
BAC T5L23 maps to near 19 cm on the Lister & Dean RI map and is
assigned to YAC C1C8B1. Position 1 of T5L23 is oriented toward the
telomere and position 116473 is oriented toward the centromere. For
more information on the mapping, sequencing and annotation of
T5L23, please see <http://www.cshl.org/Arabidopsis/T5L23-titlepage.html>.
A graphic view of our annotation is also available at this url.
Gene models are built with exons predicted by GenScan
(<http://CCR-081.mit.edu/GENSCAN.html>), MZEF
(<http://www.cshl.org/genefinder>) and GRAIL
(<http://compbio.ornl.gov/tools/index.shtml>) and with splice sites
predicted by NetPlantGene
(<http://www.cbs.dtu.dk/netgene/cbsnetpgene.html>). Genes are
numbered according to the scheme BAC.gene_number. Typically, these
numbers progress from 1 upwards as one moves from position 1 of the
BAC. Protein sequences encoded by the genes are assigned to a
functional category with the aid of similarity searches and
comparison to the Prosite
(<http://expasy.hcuge.ch/sprot/prosite.html>) and Pfam
(<http://pfam.wustl.edu/>) libraries. A description of these
categories can be found at
<http://muntjac.mips.biochem.mpg.de/arabi/>. Genomic repeats are
typically located by TBLASTX analysis and an attempt is made to
classify the function of each repeat as either transposon, putative
microsatellite, LINE, direct repeat, centromeric repeat, etc.

REMARK
COMMENT

If you have any questions or confirmatory or contradictory evidence
concerning the annotation of T5L23, please direct email to Larry
Parnell at parnell@cshl.org.
Location/Qualifiers
1. .116448
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="IV"
/map="19 cm"
/note="Hybridizing YAC is C1C8B1"
1. .31783
/note="overlap with F9H3, GenBank accession number
AF071527, from position 1 to 31783; Region: overlap with
F9H3"
complement(2025. .3495)
/gene="T5L23.1"
/note="encodes putative M-type thioredoxin; identical to
F9H3.15, GenBank accession number AF071527; genomic copy
of EST AA042639; genomic copy of EST AA650755; gene model
last edited on 5 Jan 99"
/evidence=not_experimental

FEATURES
source

1. .116448
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="IV"
/map="19 cm"
/note="Hybridizing YAC is C1C8B1"
1. .31783
/note="overlap with F9H3, GenBank accession number
AF071527, from position 1 to 31783; Region: overlap with
F9H3"
complement(2025. .3495)
/gene="T5L23.1"
/note="encodes putative M-type thioredoxin; identical to
F9H3.15, GenBank accession number AF071527; genomic copy
of EST AA042639; genomic copy of EST AA650755; gene model
last edited on 5 Jan 99"
/evidence=not_experimental

misc_feature

1. .31783
/note="overlap with F9H3, GenBank accession number
AF071527, from position 1 to 31783; Region: overlap with
F9H3"
complement(2025. .3495)
/gene="T5L23.1"
/note="encodes putative M-type thioredoxin; identical to
F9H3.15, GenBank accession number AF071527; genomic copy
of EST AA042639; genomic copy of EST AA650755; gene model
last edited on 5 Jan 99"
/evidence=not_experimental

gene

1. .31783
/note="overlap with F9H3, GenBank accession number
AF071527, from position 1 to 31783; Region: overlap with
F9H3"
complement(2025. .3495)
/gene="T5L23.1"
/note="encodes putative M-type thioredoxin; identical to
F9H3.15, GenBank accession number AF071527; genomic copy
of EST AA042639; genomic copy of EST AA650755; gene model
last edited on 5 Jan 99"
/evidence=not_experimental

CDS	complement(join(2025..2341,3252..3495)) /gene="T5L23.1" /note="similar to A. thaliana protein F21B7.7, GenBank accession number AC002560; M-type thioredoxin is also present in cyanobacteria and chloroplasts; functional catalog ID=02" /codon_start=1 /evidence=not_experimental /product="putative M-type thioredoxin" /protein_id="AADI5308.1" /db_xref="GI:4263039" /translation="MAAFTCTSRPPISLRSETRIVSSSPASSLSRRMFAVLPESSG LRLRLSPASLTSIHQPRVSRURRAVVCQAQETTDIQVNDSTWDSLVLKATGPVV DWAPWCGCKMIDPLVNDLAQHTGKIRFYKLNTDESPTQGVGRSIPITMIFV GBEKDTIIGAVPKTTLTSSLDKFLP" complement(4139..4163) /note="function=putative_microsatellite" /rpt_type=tandem /rpt_unit="a"	GenBank accession number AF071527; gene model last edited on 5 Jan 99" /evidence=not_experimental join(9482..9639,9745..9847,10690..10923) /gene="T5L23.3" /note="identical to F9H3.17, GenBank accession number AF071527; similar to F21B7.5, GenBank accession number AC002560; functional catalog ID=99" /codon_start=1 /evidence=not_experimental /product="hypothetical protein" /protein_id="AADI5310.1" /db_xref="GI:4263041" /translation="MWKLTKRIGLVRLAFAAGAAALIVMITSRRASFLAISLEA KYTDMAAFKYFVIANAVSVYSLVFLPKESLWKFVVDLWMLTLLTSSLSAALA VAQVGKGNANAGWLPICGVKPCQDITGALITAGFVALVLYVLLLYSLHAVVDFFL LOKS"	CDS
repeat_region	12953..18635 /gene="T5L23.4" /note="encodes putative glucan synthase component; identical to F9H3.18, GenBank accession number AF071527; gene model last edited on 5 Jan 99" /evidence=not_experimental join(12953..14823,14984..17006,17187..18635) /gene="T5L23.4" /note="similar to 1,3-beta glucan synthase; identical to F9H3.18, GenBank accession number AF071527; functional catalog ID=01.05.99; functional catalog ID=09.01" /codon_start=1 /evidence=not_experimental /product="putative glucan synthase component" /protein_id="AADI5311.1" /db_xref="GI:4263042" /translation="MSLRHRTVPOTGRPLAAEAVGIEEEPYNIPVNNLADHPSLR FPEVRAAAALKTVGDLRRPPYQWRSHYDLDLWALFFGQKDNVRNOREHVLHLA NAQRUSPPNDISLDSAVRFRKLLANYSWCSYLGKNSNIWISDRNPDSRREL LVYGLYLLIWGEAANLRFPPECICYIFHNNASELNKILEDCLDENTQGYLPSSLGSEN AFLTGVRKPIYDTIQAEIDESKNGTVAHCKWRNYDDINEFYFTRDFCFSKLWPLDLGS NFPSKSGVGTGTFVRRRTFFYLSRFDRLWMLALFLQAAIIVAWBEKPTSSVTR OLWNLAKRDVQVRLTVELTWSGMRLQALDAASQYPLAVSRKHFRFRLMKVIA AAWNIVAFVLYTNWQKQROWSNAATTKIYQIYAVGAFVLPEILALALFIIPW MRNPLEETNWKIFFALTWPGQSFVGRLEGLVDNIKYSTFWIPLATKFTFSYFL QVGMPIKPSKLLNKLNDVDYEHQFQSDNSRFSVALLWPLVLIYIMDQIWIYAIYSS IYGAUVGLFDHLGEIRDMGQRLRFQFASAIQFNLMPEQLLNAGFGNKFQDGIHR LKIRYGFGRPFKKLESNOVEANKFALIWNEIILAFREEDIVSDREVELLELPKNSMDV TWIRPFCFLCNELLALSOARELIDAPKMLWHKICKREYRCANVEAYDSIKHLLD SIIVDEEHSIITVFQIINQISQEQFTKTFRVDLLPKIYETLQKLVGLVNDSESTD SGVNVNVLQSLYEIATROFFIEKKTTEQLSNEGLPRDPASKLQFNALRLPDASNE PQVRVRLHTILTSRDSMHSVPYNLEARRIAFFSNSLFMMNPHAFQVEXOMAFSVLT PYSSEVWYSKEQLRNETEDGISTLYLQITIVADEWKNFKERMHREGIKTDSLELWTK LRDRLWASVYRGQTLARTVGRMMYVYRALKMLAFLDSASEMDIREGAQELGSLVRNLQG	Query Match 59.4%; Score 571.6; DB 8; Length 116448; Best Local Similarity 92.2%; Pred. No. 1.6e-79; Matches 636; Conservative 0; Mismatches 49; Indels 5; Gaps 3; QY 1 TGTGTTTCTCAGAAATAGCACGAAATATTATATAAAGCATGCAATTCCTCTTATAGATCG 60 DB 8582 TATGTTTCTCGAATAACACGATATATTATTATAAAGCATGCAATTCCTGTTATACATCG 8523 QY 61 CGAAG-TTTAAAAAACATATAGAAITGTTTACAATATTACATGGGTTTTTATTGGATAAC 119 DB 8522 CGAAGTTTTTAAAAACATATAGAAITGTTTACAATATTACATGAGTTTTTATTGGATAAC 8463 QY 120 ATGACAAATATTATTATTTTCATGAGTTTTTATTGGATAGCATGACAAATATTATAATA 179 DB 8462 ATGACAAATATTATATATTACATGAATTTTTTGTGGATAGCATGACAAATATTATA 8403 QY 180 TCAGTGTTAATAACATGTTTTTGTCTTAAATAACATGCATTTTAAAAATCAGACATTTGTT 239 DB 8402 TCAGTGTTAATAACATGTTTTTGTCTTAAATAACATGCATTTTAAAAATCAGACATTTGTT 8343 QY 240 TTTAAATCAAACTAAATCTCTTATATACAAACGACATTTGACGGAAAAATTCAGGTAAAAAG 299	
LTR	5120..6848 /note="function=LTR_transposon" /rpt_type=dispersed 5120..5124 /note="function-terminal_flanking_direct_repeat; cognate A of flanking direct repeat of LTR transposon from position 5145 to 6873; cognate B residues from position 6869 to 6873" /rpt_type=flanking 5125..5128 /note="cognate A of LTR flanking transposon from position 5145 to 6873; cognate B residues from 6665 to 6868" 5508..6632 /gene="T5L23.2" /note="encodes putative reverse transcriptase; likely functions as an element of LTR transposon from position 5145 to 6873; gene model last edited on 5 Jan 99" /evidence=not_experimental 5508..6632 /gene="T5L23.2" /note="likely element of LTR transposon from position 5145 to 6873" /codon_start=1 /evidence=not_experimental /product="putative reverse transcriptase" /protein_id="AADI5309.1" /db_xref="GI:4263040" /translation="MGRNCKSVRRSSPDLLVIASSKIHQSSGSSARAPGSHIAA IHQAQYSSRRFEVSDPNSSPHLVSSDHPGLVLAPELIDGNSYGTWIIAMTTS IEAKNLGFDGSIKPDPPDDPYCKIWRRCNSMVKSLNSVKEIYTSILYPTTAA IMKDLYTRFKSLPLRYKLQOIHSLRQGNLDLSSYHTRKQTLWBELTSLQIAPRTV EDLIERETNRVIDFLMGLNDCYDAVRSQILMKKTLPSLSEVENMIDQDEIORSARIS TTPGMTSSVFAVSNQSSQSLNGDTYQKRPVCTYCSRPGHVEDTCYKKGHGYTSPK SKQIANTELQIVDPFKLLYPHFQRLSKMGLLNIFVPS"	repeat_region	
repeat_region	5145 to 6873; cognate A residues from position 5150 to 5353" 6844..6848 /note="function-terminal_flanking_direct_repeat; cognate B of flanking direct repeat of LTR transposon from position 5145 to 6873; cognate A residues from position 5145 to 5149" /rpt_type=flanking complement(7885..7904) /note="function=putative_microsatellite" /rpt_type=tandem /rpt_unit="at"	repeat_region	
repeat_region	complement(7905..8818) /note="function=unclassified; similar to MXI10, GenBank accession number AB005248" /rpt_type=dispersed 9482..10923 /gene="T5L23.3" /note="encodes hypothetical protein; identical to F9H3.17,	repeat_region	
gene		gene	

CDS
complement(join(10335..10525,10983..11078,11590..12133,12345..12500,12627..12772,12871..13501))
/gene="AT4g03490"
/note="contains similarity to mammalian ankyrins
similarity to: similar to A. thaliana hypothetical protein F21B7.8, GenBank accession number AC002560"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB77834.1"
/db_xref="GI:7270672"
/db_xref="UniProt/TrEMBL:O9ZT77"
/translation="MGKQEFKMEENEIPVLDQVTFQGTTLHLAAIYCHDHLVRRILAYELNLRKGLNCFVPSFHYQTLVLRNKGDLALHVAAGHKLIVGLLIDCLRQPDITIMYIGSQWIGNIFRVSNNDGNTALHSLKGNHVSQLVREDRSTCFLLKEDVPLVMAAEGYVSLVEHMLRGLDASFVGSVLCVAAVKSQNDIITAVLESDDLVSREDGRTPATAAGSDYDGVQHMLTRFASSTQGVNHLVAAKSNARAVGYLLKSDVRLNEODIEGNTPLHLASSNHPKWLIMWALVAAGTTTPRVHLRADIPGLTDEDLILKHKDRVNTLLVATLVATMAFAAGSLVPLGVNTEFKNSVKSVEESAPFAPVINCSTAVYTAIVSTVALIGTQADLKCMUTTFKFIPLPLGFSIIAMSLAFVAGLYLVGHYHLAIPLVLAGSGFYLMALLLLIIPYASPTYFTLSRSLNSLVQMSKEDVDVSNOLVPAPTEELALIKAIQVNRCIQGNIGVTSLSYLLKHVAPINVLLHLKLVNHYSLVPSGHLYDGPKTYSQL"
complement(10335..10525)
/gene="AT4g03490"
/number=1
intron
complement(10526..10982)
/gene="AT4g03490"
/number=1
exon
complement(10983..11078)
/gene="AT4g03490"
/number=2
intron
complement(11079..11589)
/gene="AT4g03490"
/number=2
exon
complement(11590..12133)
/gene="AT4g03490"
/number=3
intron
complement(12134..12344)
/gene="AT4g03490"
/number=3
exon
complement(12345..12500)
/gene="AT4g03490"
/number=4
intron
complement(12501..12626)
/gene="AT4g03490"
/number=4
exon
complement(12627..12772)
/gene="AT4g03490"
/number=5
intron
complement(12773..12870)
/gene="AT4g03490"
/number=5
exon
complement(12871..13501)
/gene="AT4g03490"
/number=6
misc_feature
13514..15185
/note="marker mi233"
14170..17288
/gene="AT4g03500"
join(14170..14204,15103..15848,15939..16477,16650..17288)
/gene="AT4g03500"
/note="hypothetical protein with some similarity to ankyrin
similarity to"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB77835.1"
/db_xref="GI:7270673"
/db_xref="UniProt/TrEMBL:O9ZT78"
/translation="MKNESRTSEISKLIGISGVGFSMGPRPDVHLDRVAENHQAIPTNNVRSPLDLSLTFDSTKEMDPKTMVAARAGKENVLRNSNYSISVAPTLVNRGNTLHLAASSGHVSLVRYIIQKCPGLLKSMMGVEALHIAAAGHLDDVVNLIIDFTNDISCTNLPLVAKRIYFAKNKNQDTHALVHAKGKEVASYLVSAKSLSFVANRDGFSF

LYLAIBAGHTSLVTTMCHGTNELSSKVGGRSIVHAAKANRNDILDALLSKDASLNLKDEGRSLSFGASIGYQGSYFLFDKNRKVYFSDDDGLFPPTHMAKYGHVQLEELKHCEPAEIELLDRDQONILHAAKYGLKVKIFLSCKDKNKKLLNEODVNGTPLHLATINHPKVVSMTWDRVLDKRNKYGFTALDVAEENIDSSYIVHORTWALINA GAPKSTPTENLRSPKPDGKVKDRVNTLVALTLVATMTTGTATLGGVNSFP HIGMAVLAKRTAFQVFLVCDTLAMYSIIITVALIWAQUGDLSIILKAFNIALPFLGLALTSMSIAFMAGTYAVSHLPLGLGYFVLGIGIIFLLVLLLLLPPYSPYAHAPLRLH IFYYPYFLKLLAAGDNKNIIVDVYTASDE"
14170..14204
/gene="AT4g03500"
/number=1
intron
14205..15102
/gene="AT4g03500"
/number=1
exon
15103..15848
Query Match 59.4%; Score 571.6; DB 8; Length 159629;
Best Local Similarity 92.2%; Pred. No. 1.5e-79;
Matches 636; Conservative 0; Mismatches 49; Indels 5; Gaps 3;
QY 1 TGTGTTTCTCAGAAATAGACGAAATATTTATATAAAGACATGCAATCTCTTATAGATCG 60
DB 29859 TAATGTTCTCTGAATAACACGATATATTTATTAAGACATGCAATCTCTTATAGATCG 29880
QY 61 CGAAG-TTTAAAAAACAATAGAAATGTTTACAAATATTACATGGGTTTTTTATTGGATAAC 119
DB 29799 CGAAGTTTTTTAAAAACATATAGAAATGTTTACAAATATTACATGAGTTTTTTATTGGATAAC 29740
QY 120 ATGACAAATATTTATTTATTTTCATGAGTTTTTTATTTGGATAGCATGACAAATATTAATA 179
DB 29739 ATGACAAATATTTATTTATTTTCATGAGTTTTTTATTTGGATAGCATGACAAATATTAATA 29680
QY 180 TCAAGTGTAAATACATGTTTTTTGTTCTTAAATACATCATTTTAAAAATCAGACATTTGTT 239
DB 29679 TCAAGTGTAAATACATGTTTTTTGTTCTTAAATACATCATTTTAAAAATCAGACATTTGTT 29620
QY 240 TTAATAATCAAAATCTAATCTCTTATATACAAACGACATTTGACGAAAAATTCAGGTAATAAG 299
DB 29619 TTAATAATCAAAATCTAATCTCTTATATCACAACGACATTTGATGAAGAAATTCAGGCAATAAG 29560
QY 300 AGAAATTAAGAATGAGAGATAGAGATTTCTATGAAAAAAGAAAGAGAGACATGTAG 359
DB 29559 AGAAAAACAAAGAAATGAAAGATGAGAGATTTCTATGAAAAAAGAAAGAGAGACATGTGG 29500
QY 360 GTCAACAAAAATAAAGAGATATGATATATTTTATGAGAGGTGGTCAAGATTTATTTTAG 419
DB 29499 GTCAACAAAAATAAAGAGATATGATATATTTTATGAGAGGTGGTCAAGATTTATTTTAG 29441
QY 420 GAGAGGGAGAGAGAAATAGAAAAAGAAATGACATGGTGAATCTGAAGAGAGATGAATTTGT 479
DB 29440 GAGAGGGAGAGAGAAATGAGAAAAAGAAATGACATGGTGAATTTGAAAGAGATGAATTTGT 29381
QY 480 GTTAAAGATGAGAGAGAGAAAGAGAACTCCATGGCTAAAGTCTCGTAAAGAGATGAAAAA 539
DB 29380 GTTAAAGATGAGAGAGAGAAAGAGAACTTTCATGGCTAAAGTTCGTAAAGAGACATGAAAA 29321
QY 540 GAAACAAAAACGAGAGAGAGAAAGAGAAAGCGCTAAAAATAGACTTAATTTGCCAAAAATTT 599
DB 29320 AAAAGAA---AAAG 29264
QY 600 CTGTAGCCGACAAATACTATTGTCCTCAAGGTTATTTTGTGTATCTCTTTTGAAGTCAAAA 659
DB 29263 CTGTAGCAGACAAAGTACTATTGTCCTCAAGGTTATTTTGTGTATCTCTTTTGAAGTCAAAA 659
QY 660 GTTATTTCTTACATATATCTCTTAAAAATATA 689
DB 29203 GTTATTTGTATATATCTCTTATATATA 29174

RESULT 12
AC007203/c
LOCUS AC007203 95190 bp DNA linear PLN 30-OCT-2002
DEFINITION Arabidopsis thaliana chromosome I BAC T10P12 genomic sequence,

ACCESSION VERSION KEYWORDS SOURCE ORGANISM	complete sequence. AC007203 AC007203.3 GI:4996903 HTG. Arabidopsis thaliana Arabidopsis thaliana (thale cress)	repeat_region	122,554-129,271 of 'IGF' BAC
REFERENCE AUTHORS	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eucosids II; Brassicales; Brassicaceae; Arabidopsis. 1 (bases 1 to 95190) Fedorispiet,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F., Altati,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P., Gonzalez,A., Kremenetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S., Luros,S., Schwartz,J., Shinn,P., Toriumi,M., Vyotskaia,V., Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W. Unpublished	gene	/note="overlap with bases clone F1121, gb AC005687" 9474..10628 /rpt_family="LTR" complement(19003..23764) /gene="T10P12.1" complement(join(19003..19302,20071..20661,20728..20928, 21074..21678,21933..22099,22247..22640,22931..23131, 23362..23764)) /gene="T10P12.1" /note="Similar to reverse transcriptase; Similar to reverse transcriptase, part of putative retrotransposon"
JOURNAL REFERENCE AUTHORS	2 (bases 1 to 95190) Fedorispiet,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F., Altati,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P., Gonzalez,A., Kremenetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S., Luros,S., Schwartz,J., Shinn,P., Toriumi,M., Vyotskaia,V., Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W. Submitted (03-APR-1999) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA	CDS	/codon_start=1 /protein_id="AAD39270.1" /db_xref="GI:5080760" /translation="MDMIFINVGDVRRKIENSKTAAEAWGTDLILYLVKSLPNRVYL OLKVNYRMQNSLTLEENVDFPKMISDLNNIQIOVPDEVOAQVAILIUSALPDCSDM LKETLYKGREGIKLDDVISAATSKLELRDSTSEVNLGESKGDHGEWILIDTECSFHM TRKXYHIDFEAKSGKVRMPNNSFLEVGIGIKGRFTNQDGTSCFGNDRISKIKFCENF VVGKTRTSGTGAQHTVEKLDYVHSDLWGSSNPYSLSKQCQFFSTFDWGRKKWVIN FUNKIDEAFTFTWKWVETOSERLKKLRTDNGLESCNQKDFGFKKEEIVRHRTS PSSAVENKIPEELMTSVIPSLRGLRFGCIVVHSEQGLDKPKAKKGVPVGPSPHDS TSEENQRTQIARDRPRROVLPSRLKDYEMDEEVLDEIAGVAYLITEDGGNPEPVS FOEALQSDSGKWLAEABEIOSLIKNKTWLVVERNKSKQPIGCKWFKRAGIAGVE KRFYARLVTKGYSQKEGIYFOEIPSPVKVYSIRLLSIVFHLMDLQOMPKTAFL HGLDEMIPMDQEGYVHEKYPRIYLLYVDDILIASKKKKHCELVKLNSEFEM KOLGDAKILGMEIIRDEKGGTLTISQEGYLLKLRPATVEYVQSKNVRTVYRSVA GSLMYSMITGTPDLAHSVGLCLCMFMRKPLKDHQAVKWLIRLCIVTMDRKLVCYKNGG TNALEGVDSYAAADKEKRSTGVTFTGNTISWRSSLOKVVALLSTKAEYMALTDA TNFTWLGKLNELGFTORSVNIHCDOSVIALAKNAVYHEHTKHIDVKYHFTRELAK ANNELITKLGEGSLEEEKQVLANKKEYDDAEATYNNKLVLELKEELKIPPPVFNLIIP NSLGNRSAGLAEVGVTDHASSNLRDEDMCPSEBEPAV" 29564..30696 /rpt_family="LTR" complement(31637..35319) /gene="T10P12.2" complement(join(31637..32634,33345..33502,33679..33916, 34104..34307,34383..34539,34828..35319)) /gene="T10P12.2" /note="Similar to reverse-transcriptases; protein is flanked by LTRs, potential retrotransposon"
TITLE JOURNAL	Direct Submission Submitted (16-JUN-1999) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA	repeat_region	/codon_start=1 /protein_id="AAD39271.1" /db_xref="GI:5080761" /translation="WMEQLRIGTGYFGSGTSPEDASWRSRVERNFGSSKCSVEYL DLAVHFLGDAHLWRSVTAKRQDMFADWAFBNFNAKYCQPEALDRMEARFLELTQ GSRVREYDREPNRLLYAGQMEDDQAQMRFLRGLRDPWRCVRCSVOYATKAALVE TAAELMVAVVQPAVQOQGAOVQOQNVQQLAHIGAAPQGYTTREIGSTSRATIGL LAHEVCVGTLLVDGVEAHVLFDSGASHCFITPKSASRGNIIRGDPGEQSVKVGQGF LEVLGRKAVEAEKMEKCEAYLMTISMPESVQVAVNIRVQEFDEFQPSLQGLP PWSDFPTIVLEPGFTALLSKHPTGWLQQRWIDELLQURGATCFKIDLTSGVHQIPI ABADRKTAFRTYRGHFEFVTPFAEALDAIWIIVDLRTKSAHLAIIRKTGAAVLIT KKYVSEIVKLHGVPVSVISDRDSKFTSAFRAFMGTQVOMSTAYVHHPQTQGSRT IOTLDEMLRMVLDMGHWADOLSLIEFAYNNYQASIGMTPEVILYGRPCITQCSMT QVGRSIIYGANYVQSTETTERIGVILKLNKMEVQDRHSYADKRRELEFEVGERVYLKMA KURGNRSISETKLSFRIMGPRVIVPVAIRLELTDVMAFRHFKVHVSMURKULCH MDDEVLAKIPLEDLQNMTLEARPGRVLRRIKELRRKKIPIKVLWDCDGVTEETWEP EYRMKARF" 35970..37136 /rpt_family="LTR" 37243..42330 /gene="T10P12.3" join(37249..37347,37714..38098,38266..38532,38897..39097, 39215..39508,39603..40889,41069..42330) /gene="T10P12.3" /note="Similar to retrotransposon polyprotein"
REFERENCE AUTHORS	1..95190 /organism="Arabidopsis thaliana" /mol_type="genomic DNA" /cultivar="Columbia" /db_xref="taxon:3702" /chromosome="I" /clone="T10P12" 1..6718	gene	
FEATURES source	Location/Qualifiers	CDS	
misc_feature	1..6718	repeat_region	

OMVSRGKNNVVAARTYKSYTKPFLITSGAPVAHRSKLKLCADNNLWMLGNSVSH
CSGCKGVFLWLGVEFGNFMSELGUVVVLGVEMLETLAVPKLVEIHESDLTEGS
LSLSLAIEMEMLVVEVLESEPTIFDIHIELEPPFEGLEHSISLKGVSIIIVRSVRY
PHATNVVLEKMEVEMLVGTGIIRPTIIPDKFPIVIELLDLHGMIFSKIDLCSGYH
QURMPEDIEKTAFTLEHGYBFLVMSFGLTNAPATFQSSMNRISKKLLRKEVLFVFD
HLIIVSPSESHRLEVLQDLNKLKYNMKCSFGVQVVEYLGHIIISAEGVVTFD
TKTEAHNNQVFNLSKQLRFLGLTGYIKFVRGYSIAKLTITELLKRDKFLWLTWQ
DAFEQVQKAMILAPLVGFPDFDKFVLETFDFTGVGAVLMOBKRRITVFSHDLTPRE
OLKSAVERELMAMVMDLVKHYLIGRKEPKVHTDORSKFLLEKEVNEVYORMLTRL
LGCMDIVVYKPGVANKAADGLSRIQHPVYALLALTPVPYVIOQLDLYKIDEDEHIQS
VIKQHDNQLQSHYQVGLNRYKQLLVFPKHSFPIILQETVWEDISVDPIE
GUPTSGVNVILVVDRLSKAVYFGLGKHLFKAIDVANKFSEVVKLYGFPKSIIVSDR
DRIFLGAVKDMFKLSQTLKSTSYRPTQDQTEVINRCLERYLRCFTSSHPRAWHK
YLAWEFSNTAYHSALKTSFPKVVYGRPPDPSHLVRPELGSTONWDLVEQLKERDLME
HIRANLQAEIMKRNADKRDVFAVGWDVHLKLOPYQRVTVRRPCVKLAARFEG
PEIIVERICNAVIRINLPASSKTHNVNSQLKVEGEOOVQATNPPLKVEDEFLP
EKIVDISFTDKNGKFLVQDRGVSNSWMPYKRFVQFPQFKLEDKLIIFLARSIDM
THEAYVRQTRGQRHLURVEGEDVAVTPPSLTIRNSAKDPGFQVLUGQFSGLHPTRC
TA"

gene complement (47812..49281)
/genes="T10P12.4"
CDS complement (join(47812..47877,48046..48223,48314..48405,
48708..48828,49010..49281))
/genes="T10P12.4"

/note="Hypothetical protein"
/codon_start=1
/protein_id="AAD39273.1"
/db_xref="GI:5080763"
/translation="MAISLATAYISPCPTPESSNSPSPRTLSVRLPSQIRRRFSVQ
SPSSSTRFAPLTVSSFTNYLFRFIDAIDFETKFNVLIVLTSSSFVEANRLADGLAN
YAFSLPGLWLESPPDVSSILNADVAGVRAAKQTFNSFDLLQNSDKPLVDFYA
TWGCPQLMVPILNEVSETLKDIIVAVKIDTEKYPGLANKYQIEALPTFLFKDGLW
DRFEGALPANQLVERTENSLOVKQ"
complement (50976..52602)
/genes="T10P12.5"
CDS complement (join(50976..51522,51912..52404,52575..52602))
/genes="T10P12.5"
/note="Hypothetical protein"
/codon_start=1
/protein_id="AAD39277.1"
/db_xref="GI:5080767"

/translation="MCLAQMTITSQLPVRYLGLPLMTKAMTADHYLPLIEKIRKISS
WTRFLSYCGRLQILKSLVMSITNMFSSAFRLPGNCKEIERLCSAFWISGPDLLKTHN
AKIAWKGCLPCEGGLGRPLKEINTVCGKLINRLLASOTSLGWQVOTVLIIRNN
FWAIKASSYQGSWMCMVPOATPKFAITWLGWHNRLSTGDRMKNGWADSTCVPCQD
PLETRDLHFCHYANQIWEIIAKGPMGVQYTSNWDQLASLAGTSLEFLPCFFSGML
FRPPFTQRESEITVGMKEIQBOSTNWRKPKQYETESPFSAITREKNMRTSSVFGS
PLEANLSLTENFLFLKCCCT"
complement (60126..62065)
/genes="T10P12.6"
CDS complement (join(60126..60521,60607..60800,60875..61272,
61788..62065))

Query Match 42.3%; Score 406.8; DB 8; Length 95190;
Best Local Similarity 79.2%; Pred. No. 4.7e-54;
Matches 588; Conservative 0; Mismatches 82; Indels 72; Gaps 6;

QY	1	TCTGTTTCTCAGAAATAGCAGCAATATTTATATAAAGCATGCAATCTCTTATACATCG	60
Db	68685	TGTGTTTCTCTTAAATATCATGATATATTATATAAAGATGCAATCTCTTATAGAATA	68626
QY	61	CGAAGTTTAAAAAACAATATAGAAATTTTACATATTTACATGGTGTGTTTATTTGGATAACA	120
Db	68625	CGAAGTATTACAAATATATAGAAATTTTACAAATTAATAGGTTTATTTATTTGGATAACA	68566
QY	121	TGACAAATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	180
Db	68565	TGACAAATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	68506
QY	181	CAGTGTATAACATGTTTGTCTTAAATACATGCAATTTTAA-AATCAGACATTTGTT	239
Db	68505	CAGTGTATAACATGTTTGTCTTAAATACATGCAATTTTAAAGAAATCATACTTTGTT	68446

QY	240	TTAAAAATCAAAATCTAATCTCTTATATACAAACGACATTTGACGAAAAATTCAGGTAAAAAG	299
Db	68445	TTAAAAATCATATCTAATCTCT--TATCACAACGACAAATGATGGAGATTCAGGCAAAAAG	68388
QY	300	AGAAAAATAAAGATGAGAGATGAGAGATTTCTATGGAAGAAAAAGAGAGAGAAATGTA	358
Db	68387	AGAAAAAGAGAAATGAGAGATGAGAGATTTCTATGGAAGAAAAAGAGATGAGAAATGTA	68328
QY	359	GCTGACAAAAATAAGAGATATGATGATATTTTATGAGAGGTGCTGAGATTTATTTA	418
Db	68327	GCTGACAAAAATAAGAGATATGATGATATTTTATGAGAGGTGCTGAGAGAA-TATTTCTA	68269
QY	419	CGAGAGGGAGAGAGAAATAGAAAAAGAAATGACATGCTGAATCTCTGAA---GAAGATGAA	475
Db	68268	CGAGAGAGAGAGACAAAAGAGAGAAAAATGACATGCTGAATGCTGAAGATGAGATGAG	68209
QY	476	TTGTGTTAAAGATGAAGAGAGAGAAAGAACTCCATGGCTTAAAGTCTCGTAAAGAGATGA	535
Db	68208	TTGTGTTCAAAGATGAAGAGAGAGAAAGAACTCCATGGCTTAAAGTCTC-----68162	
QY	536	AAAAAGAAAAAGAGAGAGAAAGAAAGAAAGCTAAAAATAGACTAATCTATTGCAAAA	595
Db	68161	-----ATTACAAA 68153	
QY	596	ATTTCTGTAGCCGCAAAATACTATTGTCGCAAGGTATTTTGTGTATTTCTTTTGAAGTC	655
Db	68152	ATTTCTTACCAGACTAGTACTATTGTCGCAAGGTATTTTGTGTATTTCTTTTCAAGTC	68093
QY	656	AAAAGTATTTCTTACATATCTCTTAAATAATAGCCGATACCAATTTTTCACACATGG	715
Db	68092	AAATGTTATTTTGTATATATCTCTATTTTCTTATATATATATATATATATATATATAT	68033
QY	716	ACTTCCTTTATCCAAAAGTCA 737	
Db	68032	AGAATTTTAATTTTGAATTA 68011	

RESULT 13

LOCUS	AC007399	105223 bp	DNA	linear	PLN 26-APR-1999
DEFINITION	Arabidopsis thaliana BAC F14123 from chromosome V near 69 cm, complete sequence.				
ACCESSION	AC007399				
VERSION	AC007399.1	GI:4680765			
KEYWORDS	HTG				
SOURCE	Arabidopsis thaliana (thale cress)				
ORGANISM	Arabidopsis thaliana				
REFERENCE	Arabidopsis thaliana				
AUTHORS	Habermann, K., Nascimben, L., Vil, M.D., Matero, A., Rodriguez, M., Shah, R., Swaby, I., Shekher, M., O'Shaughnessy, A., Huang, E.N., Spiegel, L.A., Schutz, K., Parnell, L.D., Preston, R.R., See, L.H., Dedhia, N.N. and McCombie, W.R.				
TITLE	Arabidopsis thaliana BAC F14123 from chromosome V near 69 cm				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 105223)				
AUTHORS	Habermann, K., Nascimben, L., Vil, M.D., Matero, A., Rodriguez, M., Shah, R., Swaby, I., Shekher, M., O'Shaughnessy, A., Huang, E.N., Spiegel, L.A., Schutz, K., Parnell, L.D., Preston, R.R., See, L.H., Dedhia, N.N. and McCombie, W.R.				
TITLE	Direct Submission				
JOURNAL	Submitted (26-APR-1999) Lita Annenberg Hazen Genome Sequencing Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724				
COMMENT	BAC F14123 maps to near 69 cm on the Lister & Dean RI map and is assigned to YAC C1C12F8. Position 1 of F14123 is oriented toward the north telomere centromere and position 105223 is oriented toward the centromere. For more information on the mapping sequencing and annotation of F14123, please see http://www.cshl.org/Arabidopsis/F14123-titlepage.html . A graphic view of our annotation will also be available at this url. Gene models				

are built with exons predicted by GenScan (<http://CCR-081.mit.edu/GENSCAN.html>). MZEF (<http://www.csni.org/genefinder/>) and GMAIL (<http://compbio.ornl.gov/tools/index.shtml>) and with splice sites predicted by NetPlantGene (<http://www.cbs.dtu.dk/netpgene/cbsnetpgene.html>). Genes are numbered according to the scheme BAC.gene.number. Typically, these numbers progress from 1 upwards as one moves from position 1 of the BAC. Hypothetical proteins are those having no EST matches and similarity only to other hypothetical proteins; predicted proteins have EST matches but cannot be assigned a function based on current composition and nomenclature within the nr protein database. Protein sequences encoded by the genes are assigned to a functional category with the aid of similarity searches and comparison to the Prosite (<http://expasy.hcuge.ch/sprot/prosite.html>) and Pfam (<http://pfam.wustl.edu/>) libraries. A description of these categories can be found at <http://muncjac.mips.biochem.mpg.de/arabi/>. Genomic repeats are typically located by TBLASTX analysis and an attempt is made to classify the function of each repeat as either transposon, putative microsatellite, LINE, direct repeat, centromeric repeat, etc.

If you have any questions or confirmatory or contradictory evidence concerning the annotation of F14I23, please direct email to Larry Parnell at parnell@cshl.org.

FEATURES

```

source
1..105223
    /organism="Arabidopsis thaliana"
    /mol_type="genomic DNA"
    /cultiVar="Columbia"
    /db_xref="taxon:3702"
    /chromosomes="V"
    /maps="near 69 cm"

```

ORIGIN

```

Query Match      12.5%; Score 120.6; DB 8; Length 105223;
Best Local Similarity 84.9%; Pred. No. 6.9e-10;
Matches 135; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 1 TGTGTTTCTCAGAAATAGCAGCAATATTTATATAAAGCATGCAATCTCTTATAGATCG 60
Db 6381 TGTGTTTCTGAAATAGTATGATATATTTATAAATGCATGCAATCTCTTATAAACA 6440

QY 61 CGAAGTTTAAAAACATAGAAATGTTACATATTTACATGGGTTTATTGGATAACA 120
Db 6441 CGAAATATTAGCAATATATAGAAATGTTACAAAATTTACATGGGTTTATTGGATAACA 6500

QY 121 TGACAAATATTTATTTTTCATGAGTCTTTTATGGATA 159
Db 6501 TGATAAATATTACATATTTTCATGGGGTTTGTAGACA 6539

```

```

RESULT 14
166494/C
LOCUS
DEFINITION
Sequence 14 from patent US 5670367.
ACCESSION
166494
VERSION
166494.1 GI:2724471
KEYWORDS
Unknown.
SOURCE
Unknown.
ORGANISM
Unclassified.
REFERENCE
1 (bases 1 to 7218)
Dorner, F., Scheiflinger, F. and Falkner, F. Gunter.
TITLE
Recombinant fowlpox virus
JOURNAL
Patent: US 5670367-A 14 23-SEP-1997;
FEATURES
Location/Qualifiers
1..7218
    /organism="unknown"
    /mol_type="unassigned DNA"
source
10.7%; Score 103.2; DB 6; Length 7218;

```

ORIGIN

```

Query Match      10.7%; Score 103.2; DB 6; Length 7218;

```

```

Best Local Similarity 13.2%; Pred. No. 6.5e-07;
Matches 63; Conservative 240; Mismatches 173; Indels 0; Gaps 0;

QY 105 TTTTATTGGATAACATGACAAATATTTATTTTTCATGAGTTTTTATTTGGATAGCATG 164
Db 1593 TATAATTTTACATACATATATTTCTAAATATGAAGTGGTGAATTTGACATAGCGTAGCATC 1534

QY 165 ACAATATTATATATATCAGTGTTAATAACATGTTTTTGTCTTTAAATAACATGCAATTTAA 224
Db 1533 GCTTCTAGACGCATCTATTTCAGTTTTCAAAAAACGGCATGTAGGCATCACTGTAATACCT 1474

QY 225 AATCAGACATTTGTTTAAATCAATCTCTTATATCACAACGACATTTGACGGAA 284
Db 1473 ATCTATGCAAGTAGTTAAAGAGATAGAGAATTTGGTACRRRRRRRRRRRRRRRRRR 1414

QY 285 AATTCAAGTAAAAAGAGAAAAATAAGAAATGAGAGATAGAGAGATTTCTATGGAAGAA 344
Db 1413 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1354

QY 345 AGAGAGAACATGTAGTGAAACAAAATAAAGAGATATGATGATATATTTATGAGAGGTGG 404
Db 1353 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1294

QY 405 TGAAGATATTATTAGGAGAGGAGAGAGAAATAGAAAAAGAAATGACATGGTGAATCTG 464
Db 1293 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1234

QY 465 AAGAAGATGAATTTGTTTAAAGATGAAGAGAGAGAAATCCATGGCTAAAGTCTCGT 524
Db 1233 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1174

QY 525 AAAGAAGATGAAAAAGAAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 580
Db 1173 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1118

RESULT 15
BX255966
LOCUS
DEFINITION
Mouse DNA sequence from clone RP23-296F5 on chromosome 4, complete
sequence.
ACCESSION
BX255966
VERSION
BX255966.4 GI:29125152
KEYWORDS
HTG.
MUS MUSCULUS (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 19577)
REFERENCE
Andrew, R.
AUTHORS
Direct Submission
TITLE
Submitted (19-MAR-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clonerequest@sanger.ac.uk
JOURNAL
On Mar 19, 2003 this sequence version replaced gi:28974583.
COMMENT
Sequence from the Mouse Genome Sequencing Consortium whole genome
shotgun may have been used to confirm this sequence. Sequence data
from the whole genome shotgun alone has only been used where it has
a phred quality of at least 30.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
-----
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate

```

chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:

Em:, EMBL; Sw:, SWISSPROT; Tr:, TRMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-296F5 is from the RPCI-23 Mouse BAC Library

constructed by the group of Pieter de Jong.

For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACe3.6.

FEATURES

Location/Qualifiers
1..19577
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="4"
/clone="RP23-296F5"
/clone_lib="RPCI-23"

ORIGIN

Query Match 10.3%; Score 99.2; DB 10; Length 19577;
Best Local Similarity 50.6%; Pred. No. 2.1e-06;
Matches 266; Conservative 0; Mismatches 258; Indels 2; Gaps 1;
QY 57 ATCGCGAAGTTTAAAAACATATAGAAATTGTTACAAATATTACATGGGTTTTTATTGGAT 116
Db 12415 ACCACAACTTGTGTTGAATACAAATTATTCAACAAATAAATATGTGTATCCATGT 12474
QY 117 AACATGACAATATTATTATTTCATGAGCTTTTATTTGGATAGCATGACAAATATTAAT 176
Db 12475 AAGATGAGAAGAACATCTGAGTCATTTCCCAAGGAATTTGGAATTTCTCAAAGAGGATAA 12534
QY 177 ATATCAGTGTATAACATCTTTTGTCTTTTAAAA--TACATGCATTTTAAAAATCAGACAT 234
Db 12535 ATACTTTCTTTAGAAAATGATCGCGGAAAACCCAAAAAACCACAAAAACAAACAA 12594
QY 235 TTGTTTTTAAATCAATCTTAATCTCTTATATCAACGACATTTGACGGAAAAATTCAGGTA 294
Db 12595 ACAAACAAAAACAAAAACAAAAACAAAAACAAAAACAAAAACAAAAACAAAAAG 12654
QY 295 AAGAAGAAATAAAGATGAGAGATAGAGATTTCTATGGAAGAAAGAGAGAGAGACA 354
Db 12655 AAAGGAAGAGAAAGAAAG 12714
QY 355 TGTAGGTGAACAAAAATAAGAGATATGATGATATATTTTATGAGAGTGGTGAAGATTAT 414
Db 12715 AGAAAG 12774
QY 415 TTTAGGAGGGGAGAGAGAAATAGAAAAAATGACATGGTGAATCTCGAAGAAATGA 474
Db 12775 AGAAGAGAAAGAGAAAGAGAAAGAGAAAGAGAAAGAGAAAGAGAAAGAGAAAA 12834
QY 475 ATTGTGTTAAGATGAGAGAGAGAGAGAACTCCATGGCTAAAGTCTCGTAAAGAGATG 534
Db 12835 AGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAG 12894
QY 535 AAAAAAGAAACAAAGAGGAAAGAGAAAGAGAAAGAGGCTAAATAGA 580
Db 12895 AAAAAAGAAAGAAAGAGAAAGAGAAAGAGAAAGAGAAAGAGAAAGAGAAAG 12940

Search completed: November 7, 2005, 10:27:04
Job time : 4638 secs

This Page Blank (uspto)